

US-10-463-957-5

Query Match 20.0%; Score 1734; DB 12; Length 420;
Best Local Similarity 78.3%; Pred. No. 1.7e-100;
Matches 329; Conservative 48; Mismatches 43; Indels 0; Gaps 0;

QY 1 MNLVNRKQLEKMANVRFRQDEYVAILDALEBYHNMSENTEVVEKYLKLDINSLTDIYI 60
DB 1 MSLVNRKQLEKMANVRFRQDEYVAILDALEBYHNMSENTEVVEKYLKLDINSLTDIYI 60

QY 61 NTYKSGRNKALKKFKFYLTMVELELKNNSLTPVEKNLHFWIGGINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKFYLTMVELELKNNSLTPVEKNLHFWIGGINDTAINYNQWKD 120

QY 121 VNSDVTYKVFYDSNAFLINTLTKTIVESAINTNTLESFRENLDNPEFDYK 180
DB 121 VNSDVTYKVFYDSNAFLINTLTKTIVESAINTNTLESFRENLDNPEFDYK 180

QY 181 DKQGHFIDYKQIENPEFIIDNIITKTVLSNEYSKDLFALNKYIEESLNKITTANGNDI 240
DB 181 DKQGFNYKYAQREENPELIIDDIKTVLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240

QY 241 RNLEKFADEDLVLYNQELVERWNLAASDILRISMLKEDGGVLDVLDLPGIOPDLFKS 300
DB 241 RNPEPKNGESFNLYEQELVERWNLAASDILRISALKETGGVLDVLDVLDLPGIOPDLFES 300

QY 301 INKPDSTINTSWEMKLEAMTKYKEYTPGYTSKNFMDLDEVOQSPESALSSKSKSEIF 360
DB 301 IEKPSVTVDVFWEMTKLEAMTKYKEYTPGYTSKNFMDLDEVOQSPESALSSKSKSEIF 360

QY 361 LPDIDDIKVPLEVKIAFNANSVNTQALISLKDVCSDLVNQIKRKYIINDNLNPSINE 420
DB 361 SSLGDMEASPLEVKIAFNANSVNTQALISLKDVCSDLVNQIKRKYIINDNLNPSINE 420

RESULT 13

US-10-463-957-7

Sequence 7, Application US/10463957
Publication No. US20040028705A1
GENERAL INFORMATION:
APPLICANT: Ballard, Jimmy D.
APPLICANT: Spyras, Lea M.
TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
FILE REFERENCE: 5820.637
CURRENT APPLICATION NUMBER: US/10/463.957
CURRENT FILING DATE: 2003-06-17
PRIOR APPLICATION NUMBER: 60/389,685
PRIOR FILING DATE: 2002-06-17
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 170
TYPE: PRT
ORGANISM: Clostridium difficile

US-10-463-957-7

Query Match 8.9%; Score 769; DB 12; Length 170;
Best Local Similarity 85.3%; Pred. No. 1.5e-40;
Matches 145; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 MNLVNRKQLEKMANVRFRQDEYVAILDALEBYHNMSENTEVVEKYLKLDINSLTDIYI 60
DB 1 MSLVNRKQLEKMANVRFRQDEYVAILDALEBYHNMSENTEVVEKYLKLDINSLTDIYI 60

QY 61 NTYKSGRNKALKKFKFYLTMVELELKNNSLTPVEKNLHFWIGGINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKFYLTMVELELKNNSLTPVEKNLHFWIGGINDTAINYNQWKD 120

QY 121 VNSDVTYKVFYDSNAFLINTLTKTIVESAINTNTLESFRENLDNPEFDYK 170
DB 121 VNSDVTYKVFYDSNAFLINTLTKTIVESAINTNTLESFRENLDNPEFDYK 170

RESULT 14

US-10-114-170-257

Sequence 257, Application US/10114170
Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114.170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 257:
SEQUENCE CHARACTERISTICS:
LENGTH: 3169 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 257:
US-10-114-170-257

Query Match 5.9%; Score 512.5; DB 14; Length 3169;
Best Local Similarity 20.2%; Pred. No. 1.6e-22;
Matches 422; Conservative 333; Mismatches 720; Indels 619; Gaps 104;

QY 14 YVKFRI--QEDEYVAILNAL--EYHNMSESSVVEKYLKLDINNTDNLNTYKKGSR 68
DB 160 YIKRTRGAEDQTITITQSLIINELLNGVDRTI--PFQKISELDIHSVENMGIKSR 218
QY 69 N--KALKKFKFYLTMVELELKNNS-----LTPVEKNLHFWIGQ 106
DB 219 KGIELVKGQELLSSILINDNKGKQISDNASKIINLLGIEYQSHKVDIEPFIHAVVAGA 278
QY 107 INDNTAINYNQWKDVSNTYTKVFYDSNAF-----LINTLTKT----- 144

Db 279 PPDNTPSYTAFLNTYKDYTYLLWIDPNAFAAKAFSGILKNIAANNVAINMRLRTPHPLAE 338
Qy 145 -----IVESATNNTLE--SFRENLDPEFY-----NKFYKRMELIYDQKHFI 187
Db 339 EMNEVILKIQIONQETIEFKETRELRKLENNRYKSLTSETKEFNFFVLESIMGMQDNYP 398
Qy 188 DYKSOIEENPEFI-----DNIUK--TYLSNEYSKDLBALNKYIEESLNKKTANNGN-- 238
Db 399 TYCISNGISNTDDISRLDFUTNVLKLSPEVONDFKSTVEKVKEDIDLLKNTISQKGFDRF 458
Qy 239 ---DIRNLEKFADEBLVRLYNQELVERNLAASDILRISMLKEDGGVYLDVVDILPGIQP 295
Db 459 QLRDINTLESFKPKQDFFVQOEMLAWNAASDQVRINILKEYGGIYVTDILPAYSD 518
Qy 296 DLFKSNKPSITNTSNEMIKLEAIMKY--EYIPG--YTSKPNFMDLDEBQVSFSAIS 351
Db 519 KVSQIINE-XSDSKRFFEDLKLRIISSEILSIKGEKYSIKG-DGLDETTLNQLNNIL-- 575
Qy 352 SKSDKSEIFLPLDDIKVSPLEVKIAPANNVI-----NOALLSLKSDSYC 395
Db 576 SEIEK-----LTDIDY-FKPVETKVVRDTFKIPRYQKQWENTWNIRGNFMTLHGSKC 630
Qy 396 SDLVINOIKRYKILNDLNAPSINEGTDFTNTMKIFSDKLASISNE-----DNM 444
Db 631 IDFILSGKKQYLEL-QRIRDNISYNNLFYTT-----EDLKSNNVAIGGIPAKKYLEHG 684
Qy 445 MFMIKITNLYKVGPAPOVRSTINSGP----- 471
Db 685 LF-----SEYRQDGTIPYVSTLNSGPMIMROMKMYKYSIGRIGEVHIXDNKLSDVNFI 740
Qy 472 GYVYGAVQDLMF---KDNSTNIHLBP-----LNFEPFKTKISQLT-----EQEITS 518
Db 741 GYVASSKDNKSFNLNVPVSGINDITPDDSSWAVRNNDINKILFEKINCHVPEKLUPTS 800
Qy 519 LWSFNQARAKSQFBEYKKGYPEGALGEDDNDLPAQNTVLDKDYVSKKILSMKTRKEYI 578
Db 801 LY-----YELDSRFFQW-----DNKSIKHVTEINKDLI--KOINLLITSSNIDV 844
Qy 579 HYIVOLQDKISYEASCNLFKDPYSILYOKNIEGSETAYYYYVADAIEKIDKXIPY 638
Db 845 KLLIKL--DRELYAISXSI-----DNPLALRSIRTLQLQANVYTSNTEPENTINFIY 896
Qy 639 QISNKN-----IKLTFIGHKSEFNTDTTANLDVDS-----LSSEIET 677
Db 897 DFKKQDILLASAIIK-----FSRNDADTKIIWYNSYMEKNVFLREVIS 942
Qy 678 ILNLAKADLSPKYLEINLLGCMFYSISABETYPGKLLKIKDRVSELMPSISQDSI-- 735
Db 943 VLRSKKVD---SYINEN-----KXNLSKEDA--GALRDYAKUKOKELFSLMLDDGGYK 990
Qy 736 TVSANQYVRINEEGKREILDHSGKWINKESIIKDISKEYISFNP-----K 783
Db 991 IITWAY---IKERDKL-----SGIYNIENSIISGHESFIIIRSNQHEWGDLSVQFK 1042
Qy 784 ENKIIVKSKYLHELSTLQEIERNANSDDILEKKVMUTECEINVAS-----NTDRQIVE 838
Db 1043 KFEFVYKSE-LSSAKSIPDDIKNK-YITDPEKRVNLVHQLDSIDKIERIAFLDISHAYP 1100
Qy 839 GRIBEXONLTSDSINYKNEFKLIESI-----SDSLYDLKHQNL 878
Db 1101 GSILLEKLQLS-----GVFSDINIIAYLLASYGVSGHSHGVVYPAPSDKLELLRHTK 1156
Qy 879 DDSHITSPEIDSKTENGPRIRINKETGNSI---FIETEKEIFSYATHSKETSNIKDT 935
Db 1157 SNSEW-I-EKITP-----YVVDILSNVSNLPPPLSEBQKILNDIKLEISKVS---EQ 1207
Qy 936 IFDNVNGKLVKVNLDAAHEVNTLSAFTIQLIEYNTTK-----ESLGNLSV----- 983
Db 1208 YPKLUEQKSSVIGIKYVDFORYENLFLSLPINQNLTPMRYFEMVLVHIGIEN 1267
Qy 984 -AMKVOVYACLSTGNTTDDASKVELV-----STAL-DETIDL 1021

Db 1268 KANREFIYKFSNLNDFLINDERVNLGLIKKYKVLSEIHRILTNTSTSPADISIPL 1327
Qy 1022 LPTLSEGLPIIATII-----DGVSLGAIAK-----ELSETNDPLL 1057
Db 1328 LQTI---CPSITTIKKTEYVGHOLTNAWTVASVVKPYDFSNLGAINSIDKSVSDVPALH 1384
Qy 1058 QEIE-AKIGIMAVNLTAATAIVTSAL-----GIASGSILL 1093
Db 1385 TIVEQAKYNLLSMWDFYNTHASIMWDTIARQHKSTNIEFHPQSLLFDRDSKGLGUSLLY 1444
Qy 1094 VPLAGISAGISLVNN---ELILODKATKVIDYFKHISLAETEGATFLDDDKIIMPQDD 1149
Db 1445 LDTGGYGGYQKLRHNDITASTLYQTK-----YNDNLKLSNRDDFLRKTQRIITWSNE 1498
Qy 1150 LVLSEIDNNNSITLGKCEIWRAGEGSGHTLTDI--DHFFSPSSITTYRKPWLSIYDVLN 1207
Db 1499 L-----GNRELKNAQLEVLELK--DPILTEGLYQRRISLITTEYHSLAQISSL 1548
Qy 1208 IKK-----EKIDFSKOLMVLNANRVFGYEMGWTPGFRSLDNDGTKLLD 1252
Db 1549 FWRVTDPNFGCHDFHSLAQALTFIKNI-----TSNR-----NFSLSYSGSI 1589
Qy 1253 RIRDHY-EGQFYWYFAFIADALITKLKPRYEDTNVRINLDG--NTRSEIYVPIVITTEQI 1308
Db 1590 -VKIYFSESLANWKYIKLPLVQTGSLURDIYLTPEKLSSTGGSINIMGHLVPV 1642
Qy 1309 RKNLSYSP-YSGSGSYSLSPYNNIDNLNVENDTWIDVNVNXXNITTESDRIQ 1363
Db 1643 ---STIYDIGV-----INGNRISST---DVKNKIRSLKINGDILQHYIN 1682
Qy 1364 ---KGLIENILSKINIEDNKI-----ILN--NHTINPY- 1392
Db 1683 THYLSEEQTKIKDIVDFLGIQDNTIKVLESIDKIPSEIOQPLHSILSRQKHEVKNLLS 1742
Qy 1393 GOINE-SNRF-----ISLTFSEILEDINI-----IIEI---DLVSKSYKILLSGNCMKL 1436
Db 1743 GULDBFSNKLKQGLSKLXTNLVSNVNFESKINSSTVEVTVDLQGRLYRVDIDTRVIGL 1802
Qy 1437 I-ENSSDIQKIDHIGFNGEHQKYPYSYIDNETKNGFYDYSK---KEGLFTAEPFN 1490
Db 1803 TKFEGINSUSEALEHNN-----IDAINSVIGLVQYARMIKNDNISAIIDAG 1849
Qy 1491 E-SIIRNIYMPDSNLIYSSKDLKIRIINKGVK-LIIGNYFKDDMKVLSLFTIEDTN 1548
Db 1850 AVSDIKNIV-----DKFLGGILTLNRYNVPFGVSGASLEGFISGLEVCAS----- 1897
Qy 1549 TIKLNGV---YLDENGVAQILKFMNNAKSALNTSNLNFLESINIKFIYFNLD----- 1600
Db 1898 --RMGGTAGRYL--SNVAKVIKL-----PLLDI--GINIWSLYDSSLNEAKAT 1939
Qy 1601 PNIEFI---LDTNF-IISGNSISIQFE-----LICDKKNIOQYFIFNEKIKETSITLYV-- 1650
Db 1940 TQIEYISTAIDVSFSSINTALSIGAIVPPLAIA-----IVPITI-FSHEVKNYAVYVNO 1993
Qy 1651 -GNRQNLIVEPSVHLDD-----SGNISSTVNFQKLYG---IDRYVKNVII 1694
Db 1994 INERHKLWEAEKYLIDNGSAKVLKINKATGIIDLNNQVLGNLYLDMRENPPIL 2047

RESULT 15
US-10-282-122A-76865
; Sequence 76865, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant

```

; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78665
; LENGTH: 4688
; TYPE: PRT
; ORGANISM: Ureaplasma urealyticum
; US-10-282-122A-76865

Query March 5.08; Score 431.5; DB 12; Length 4688;
Best Local Similarity 19.18; Pred. No. 3.4e-17;
Matches 420; Conservative 329; Mismatches 743; Indels 709; Gaps 103;

QY 13 VYVFRQDEBYVAILNALE-EYHNMSESSWEKYLKLDINLTNYLNT---YKSGR 68
D 403 LYEVIGAKKNEVDVNLKQIPYHRS-----INLNSALNTPQYTYNGD 449
QY 69 NKALKKFEKYLME-----VLELKNS-----LTPVEKNLHPFIWGGQINDTAIYNQWKD 120
D 450 INLIKAVPYVNVQVYGFQDQNNQEHQILAKVKDGTAFDQALNNNSYSLDKIVS 509
QY 121 VNSDYTVKVDNSAFNLTKTIVESATNTLESFR-----ENL----- 161
D 510 VSNPQNLV---SNFDLTSKQKLIKPAANASVDSKTKQILENLNLDLINOQKLVATFV 566
QY 162 --NDPEFDYKFKRVEIIVDKO---KHFDYVYKSOIEN----- 197
D 567 DNNDKEYKVVANVDQNNKIIFSDSNDLPKGYI-YHLAKVNNNDLKNVNLNDFELDKIID 625
QY 198 -----PEFIDN-----IKTYLSNEYSKOL--EALNKY-----IBESLN 230
D 626 KRDNLNLDSDHFDYDNDGNLEIHTQLANDLDDLKQALNNANVKCIIVDQDQIEHID 685
QY 231 KITANNNGDIRNLEKFADEDLVR--LYN-QELVERWNLAASDILRISMLKEDGGVYLDV 287
D 686 VSDANGKVIPTKRLANNPTKPNIVTLKKVVLKQNNQPNIDLISEEQSGDNHISFK- 744
QY 288 DILPGIDPLFKSNKPD-----SITNTSWEMIKLEAMKYKEYIPGYTSKNFDMLEBEVQ 343
D 745 -----KPTITAKTKENDDYEISFSPNSLANKKIKLFTKDD-----NNTNTKTVEAS 791
QY 344 RSPESALSCKSKSEIFLIP-----LDDIKV-----SPL----- 371
D 792 IGLDKALFKTSDDAIFAPDHKYTLTKIEDNKKVANIIDEISPLDRIVNKKQKGNVWADN 851
QY 372 --EVKIAPANNSVI-----NQALISLSDSYCSOLVINOIKNRY-----KI-- 409

```

```

Db 852 KHEFKIPDQKNKOLTAVYKDKNNNEIHVPKTDQDKGVIVPNPNNLFDPNKIYDFDKIVD 911
QY 410 LNDNLNPSINEGDTFNTMKIFSD-----KLASISN--EDNMFMKIIITNVLKVG 457
D 912 LNEYPNTILDRNSINKDVSAINDGVQDQARLVKAPAVSNVTVAINFQVNLFDNLKLS 971
QY 458 ---FAPDVRSTINLSGPGVYTGAYQ-----DLMFQDN 487
D 972 YNRQFALTIKAVNDLNDTKVIATYDPTNNYKLNFDTHLDANTKYKVVVDVLELNNK 1031
QY 488 STNHLLPELNRNFEFPFKTKISQTEQETISLSFNCARAKSQ-----FEYKKGVEG 541
D 1032 EKPIKLIKDDVLNFEFTSSATTIN---PPIWTKFDVVVVTNNNDTITFEIDDK---DN 1084
QY 542 ALGEDDNLDPQNTVLKDVYKKSILSSMKTRKEYIHYIVQLQGDKISYEASCNLPSKD 601
D 1085 ILKNDQKIT-YAQLALMDDDDLDDTVINPLVNFST-----NKIASINGLDDLKNS 1133
QY 602 PYSSILYQKNIEGSETAYVYVADAIEK-----EIDKYRIPYQISKNRKNIKLFFIGHGK 655
D 1134 KYSI-----KNL-----YYLNDQNKVYLFKNDVTKYE-QHFITNPKINLSF--NK 1177
QY 656 SEFNTDTF---ANL-----DVDSLSEIEFETILNLAKADISPKY 690
D 1178 SAVEQDIFADHANLFDYKDYDQKLINEDVKIYQINIDNTKNELO---IGYKV-VANNK 1234
QY 691 IEINLGCWFYSISABEETPGKLLKIKDRVSELMPS-----ISQDSITV 737
D 1235 IKFNLVG-----LKEKTYTVIKLEBALNKSASSIVNSEFDDLDPTNFTSKNTTLV 1287
QY 738 SANQY-----EVRINEEGKEILDHSGKI--NKEESIIDKISSEY 777
D 1288 GLNSIDVNNNDQTFIINAKINIGDDFQDNQOVKLIYVSDNKEIKSKAVTILKQRYNQF 1347
QY 778 ISFNPKNKIIVKSKY-----LHELSTLLOEI---RNA-----NSSDIDLEKVMYL 821
D 1348 EFSNLIKRLVTFSKIVYETNNQTLHKLDLTTHQFSINPNSNAVSLKNTNIEITKRLIV 1407
QY 822 TECEINVASINDROIVEGRIB--EAKNL--TSDSINYIKNEFKLIESDSLYDLKHQNG 877
D 1408 -----NNDQSLISAKIEVDIDNVLNTADKFNIVYQ---LENSDANNLKLATN 1454
QY 878 L-----DDSHISFEDI-----SKTENGFRIRFINKETGSIETETKE 916
D 1455 AQVYNNKKFLKFDLVSCLKINQNVYKIEISFNKPTNAY-FNFTNNKTNIVSYDQDN 1513
QY 917 IFSEVATHISKEISNIKDTIFDNV-----GKLVKVNLDAAHEVNTL----- 959
D 1514 -----KISLNNINPTSYKPKQTINKNDKSVNIDVDLQVDKQLLANQVLR 1557
QY 960 -----NSAFFIOSLIEYNTTKES--LSNLSVAMKVQVYAQLFS-TGLNTITDASKVV 1008
D 1558 LKALKQNDKNTVTDPIFLFNNNAKISPKLSNL-----IHNAYELEGVYFDDQNSVN 1610
QY 1009 ELVSTALDETDLLPTLSEGLPIATIIDGVSGLGNAIKELS-----ETNDPLARQ 1058
D 1611 DMTNNQISFNKSIHRKPIEFEPESLTTI--NYDTNNAIKTVSAHNAQVHFUKLTNDDEALN 1668
QY 1059 E-----IEAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGIS 1100
D 1669 DQIVBAVFAPTNNLNDQKVVEAKNNVNSFNEGELEFNLSGLREETIYRLIKVTFKNKP 1728
QY 1101 AGIPSILVNN-----ELIQDKATKVIDYFKHISLAETEGFTLLDD----- 1141
D 1729 NKAYELLKNGVIFPKNGSQAYEFTQKFEHKVIDVSVSSTNTTQOEITVKIDGIQRA 1788
QY 1142 ----KIIMPQDVLVSE-----IDFNNSI-----TLGKCEI 1169
D 1789 MNNKLELVYESNILGDPBEIKTTVDNNKNSVHLSFDKKEYNLVNLNLPKFGRRYSLLKINI 1848
QY 1170 WRAEGSGHGLTDDI--DHFF-----SSPSITYRKPLWSIYVZLNKKKIDFS 1216

```

Db 1849 KEVNDGQHEFVKFVNNSFDVNLQSEITASSVEEINDRAP-----DKLNQTTIKINLK 1903
QY 1217 KDLVLPNAPNRVFGYEMGTGPFGRSLDNDGTLKLDRI-RDHYEGQFYWRYFAPFIADALI 1275
Db 1904 DNDILK-----TNDIATITYDNEQKVDALVKTNAQNKY-----LEALI 1943
QY 1276 TKLKFYEDTNVRINLDGNTSFIVPVITTEQIRKNSLSYFGSGSYSLSLSPYNNVI- 1334
Db 1944 TNL-----VFNKQWIIKKIEFKNLSQTE-----IKVGKNTNVI 1977
QY 1335 ----DLN--LVENDTWI-----DVD---NVVKI-----ITIESDEIQGELIENILSKIN 1376
Db 1978 YDESNLKLIINNDPQIIGPLSTDVNSTQNIIVANNKHVISTLDFKYNPHISKNLKFKLK 2037
QY 1377 IE-----DNKIIINN--HTINFYGDINESNRFSLT-PSILEDINIIIEIDL 1420
Db 2038 FQNINGEVVYGPILTNSIIIVNNKNVINFTDNLKSNQLRLVDVYIIDNN-----DT 2092
QY 1421 VSKSVKILLSGNCKMLIE-----NSSDIQQKIDH--- 1449
Db 2093 INDKKVPKANNVTIIDIAPGKTTISKSNNTWNTTSSQFEFVINSDDGNEVLDNLEA 2152
QY 1450 -IGFNGEHQKYPYIYNETKNGFIDYSKKEGLTAEFSGESIIRNIYMPDSNNLFIY 1508
Db 2153 TISFKKGOTLLTPVKV--NIIKQNN--KYLKIGQITNLEPENRYVLESILLAKFN----- 2203
QY 1509 SSKDLKDRIINKGDV--KLLIGNYFKDDMKVUSLFTIEDNTI--KLVGVYLDENGVAQ 1564
Db 2204 KTKPLVVEILNKDDISFQOAGNYKVIQIKSQNPSTVDTKQRIKLDGIGQNAWNEKQL 2263
QY 1565 ILKEVNNAKSALNTSNNMFLESINIKNIF--YNNLDPN-----TEFILDN----- 1610
Db 2264 EITISAND-----NSTKTAIIKLEKVKLEYEFELTNLEKQRTYTKIELINDNNTKTPFN 2319
QY 1611 -----FIISGSNSIGQFEL--ICDKDN--IQPYFINFKIKETSYTYVGNRQNLIVE 1659
Db 2320 KSDSIQDKFIVLSNNQVGVGNIIEIQDRDVNHLNSAKIRFELNDLDNVL--SNDEQATIT 2377
QY 1660 PSYHLDDSGNISSTVINFSQKLYGI-----DRYVNVK 1692
Db 2378 YNNNOTTSKV---ITDQNKYLEATFSNLVINKDTIINKI 2415

Search completed: April 1, 2004, 16:55:00
Job time : 84.875 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:39:30 ; Search time 25.625 Seconds
(without alignments)
3424.942 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700
Perfect score: 8675
Sequence: 1 NMLVNKAQLQKMWYKFR1Q.....VLYGIDRYVNVKIAPNLYT 1700

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A COMB pep:*
2: /cgn2_6/prodata/2/iaa/5B COMB pep:*
3: /cgn2_6/prodata/2/iaa/5A COMB pep:*
4: /cgn2_6/prodata/2/iaa/5B COMB pep:*
5: /cgn2_6/prodata/2/iaa/5A COMB pep:*
6: /cgn2_6/prodata/2/iaa/5B COMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6759.5	77.9	2366	1	US-08-480-604A-10
2	6759.5	77.9	2366	2	US-08-405-496A-10
3	6759.5	77.9	2366	3	US-08-915-136-10
4	6759.5	77.9	2366	4	US-08-957-310-10
5	6759.5	77.9	2366	4	US-10-011-366-10
6	6759.5	77.9	2366	4	US-09-084-517-10
7	4467	51.5	2710	1	US-08-480-604A-6
8	4467	51.5	2710	2	US-08-405-496A-6
9	4467	51.5	2710	3	US-08-915-136-6
10	4467	51.5	2710	4	US-08-957-310-6
11	4467	51.5	2710	4	US-10-011-366-6
12	4467	51.5	2710	4	US-09-084-517-6
13	512.5	5.9	3169	4	US-09-453-702B-257
14	347	4.0	10182	4	US-08-134-001C-3159
15	334	3.9	2807	4	US-09-543-681A-4980
16	315	3.6	2777	4	US-09-543-681A-6124
17	304	3.5	2184	4	US-09-417-485D-6
18	289	3.3	3433	4	US-09-091-501B-10
19	279	3.2	3696	4	US-09-134-001C-5080
20	266.5	3.1	2662	4	US-09-595-694B-31
21	265.5	3.1	3248	1	US-08-353-700-1
22	265.5	3.1	3248	5	PCT-US95-16216-1
23	263	3.0	2504	4	US-09-328-352-5821
24	259	3.0	2285	4	US-09-308-375-2
25	254.5	2.9	3135	1	US-08-323-170B-2
26	254.5	2.9	3135	4	US-08-954-441-2
27	253.5	2.9	2391	2	US-08-446-855A-2

ALIGNMENTS

RESULT 1
US-08-480-604A-10
; Sequence 10, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480.604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

Sequence 2, Appli
Sequence 1, Appli
Sequence 28, Appl
Sequence 28, Appl
Sequence 28, Appl
Sequence 28, Appl
Sequence 27, Appl
Sequence 1, Appli
Sequence 8, Appli
Sequence 1, Appli
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 50, Appl

28 253.5 2.9 2391 3 US-09-150-741-2
29 252 2.9 2954 4 US-09-150-867-1
30 250.5 2.9 1296 1 US-08-480-604A-28
31 250.5 2.9 1296 2 US-08-405-496A-28
32 250.5 2.9 1296 3 US-08-915-136-28
33 250.5 2.9 1296 4 US-09-084-517-28
34 248 2.9 4536 4 US-09-180-422B-27
35 248 2.9 4583 4 US-09-108-006C-1
36 246 2.8 1786 3 US-08-973-462-8
37 245 2.8 4536 4 US-09-079-030-1
38 239.5 2.8 1346 1 US-08-471-033-23
39 239.5 2.8 1346 2 US-08-471-044-23
40 239.5 2.8 1346 2 US-08-463-483A-23
41 239.5 2.8 1346 2 US-08-471-046A-23
42 239.5 2.8 1346 2 US-08-470-566B-23
43 239.5 2.8 1346 2 US-08-469-334-23
44 239.5 2.8 1346 3 US-09-300-529-23
45 234.5 2.7 1338 1 US-08-471-033-50

```

; REFERENCE/DOCKET NUMBER: OPBD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-10

Query Match 77.9%; Score 6759.5; DB 1; Length 2366;
Best Local Similarity 76.6%; Pred. No. 0;
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;

QY 1 MNLVNAQQLQKVVYKFIQDEYVAIILNLEEYHNMSESSVVEKYLKLDINNLTDNYL 60
DB 1 MSLNVRKQLEKMANVRFTQDEYVAIILDLLEEYHNMSENIVVEKYLKLDINSLTDIYI 60
QY 61 NNYKSGRNKALKKPKVEYLTWEVLELQNSLTPVEKMLHFTWIGQINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKPKVEYLTWEVLELQNSLTPVEKMLHFTWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVFDSNAFLINTLKKTIVESATNTLTLESFRENLDNPFYKPKYKEMEIY 180
DB 121 VNSDYTVKVFDSNAFLINTLKKTIVESATNTLTLESFRENLDNPFYKPKYKEMEIY 180
QY 181 DKQKHFIDYKQIENPEFIIDNIITKYLNEYSKOLEALNKYIEESLNKI TANNNDI 240
DB 181 DKQKHFIDYKQIENPEFIIDNIITKYLNEYSKOLEALNKYIEESLNKI TANNNDI 240
QY 241 RNLKFADEDLVRNLQELVERWNLAAASDILRSMKLDGGVYLDVDILPGIDPLFKS 300
DB 241 RNLKFADEDLVRNLQELVERWNLAAASDILRSMKLDGGVYLDVDILPGIDPLFKS 300
QY 301 INKPSITNTWEMTKLEAKMYKYEIYGYTSKNPMDLDEEVQSFESALSKSKDKSEIF 360
DB 301 IEKPSITNTWEMTKLEAKMYKYEIYGYTSKNPMDLDEEVQSFESALSKSKDKSEIF 360
QY 361 LPDIDIKYSPLEKAFANNSVINGALSLKDYSCDLVINOIKRYKILNDNLNPSINE 420
DB 361 LPDIDIKYSPLEKAFANNSVINGALSLKDYSCDLVINOIKRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKIFDKLASINEDNMFMKITNYLVKGFAPDVRSTINLSGCVYTGAYQD 480
DB 421 GTDFNTMTKIFDKLASINEDNMFMKITNYLVKGFAPDVRSTINLSGCVYTGAYQD 480
QY 481 LLMFKNSTNIHLPELPELNPFPPTKISQLEQBITSLMSQARAKSQFEYKKGIFYE 540
DB 481 LLMFKNSTNIHLPELPELNPFPPTKISQLEQBITSLMSQARAKSQFEYKKGIFYE 540
QY 541 GAGEDDNLDFQNTVLDKDYYSKILSSMKTRNKEYIHYIYVQLQDKISYEASCNLFK 600
DB 541 GAGEDDNLDFQNTVLDKDYYSKILSSMKTRNKEYIHYIYVQLQDKISYEASCNLFK 600
QY 601 DPYSSITLYOKNTBEGSTAYVYVADAEIKYIPYOISNKRNIKLTFFIGHGKSEFNT 660
DB 601 DPYSSITLYOKNTBEGSTAYVYVADAEIKYIPYOISNKRNIKLTFFIGHGKSEFNT 660
QY 661 DTPANLDVDSLSSEIETILNLAKADISPKYIIBINLLGCMNFYSISABETYPKLLKIK 720
DB 661 DTPANLDVDSLSSEIETILNLAKADISPKYIIBINLLGCMNFYSISABETYPKLLKIK 720
QY 721 DRYSELMPSISQDSITVANCYEVRIINSEGRBLDHSCKWINKWESIIKDISSEYISF 780
DB 721 DRYSELMPSISQDSITVANCYEVRIINSEGRBLDHSCKWINKWESIIKDISSEYISF 780
QY 781 NPENKTIYKSVKYLHELSTLQIRNNVNSSDIDLEKKWMLTECEINVASNIDROIYVEGR 840
DB 781 NPENKTIYKSVKYLHELSTLQIRNNVNSSDIDLEKKWMLTECEINVASNIDROIYVEGR 840
QY 841 IBEAKNLTSDSINVIKNEFKLIESIGDSLVDLKHQGLDSDSHFISPDISKTENGFRIRF 900

```

```

DB 841 IEEAKNLTSDSINVIKNEFKLIESIGDALCDLKOQNELEDHSFISFEDISTDGFIRF 900
QY 901 INKETGNSIFETETKEIFSEYATHISKEISNIKOTIDFNVNGKLVKKNVLDAAEVNTLN 960
DB 901 INKETGESIFVETETKEIFSEYANHTIEISIKIKGTIFDVTNGKLVKKNVLTTEVNTLN 960
QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAFIIQSLIEYNSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDETID 1020
QY 1021 LLPTLSEGLPIATIIDOVLSGAALKELSETNDPLLRQIEIAKIGIMAVNLTAASTAVT 1080
DB 1021 LLPTLSEGLPIATIIDOVLSGAALKELSETNDPLLRQIEIAKIGIMAVNLTAASTAVT 1080
QY 1081 SALGTASGFSILLVPLAGISAGISPLVNNELILOKATKVIDYPGHISLAETEGFTLLD 1140
DB 1081 SSLGTASGFSILLVPLAGISAGISPLVNNELVLRDKATKVVDYFKHSLVTEGVFTLLD 1140
QY 1141 DKIMPODDLVLSEIDFNNNSTITLGCBIWRAEGSGHTLTDIDHFFSSPITYRKPL 1200
DB 1141 DKIMPODDLVLSEIDFNNNSTIVLGKCEIWRMEGSGHTVTDIDHFFSAFITYREPHL 1200
QY 1201 SIYDVNLNKKKIDPFSKOLMVLNPNAPRVFVEMGWTGPFRLSDNDGTYKLDRIIDHYEG 1260
DB 1201 SIYDVLEVQKEBLDLSKOLMVLNPNAPRVFAWETGWTGTLASLENDGTKLDRIIDRYEG 1260
QY 1261 QYWRYPAPIADALITKLPRIYEDTNVRLDGNTRSFIVPVTITEQIRKLSYFYSG 1320
DB 1261 EFWYRYPAPIADALITKLPRIYEDTNVRLDGNTRSFIVPVTITEQIRKLSYFYSG 1320
QY 1321 GVSLSLSYNNMIDNLVENDTWIDVNVVKNITTESDEIOKGELIENILSKLINEDN 1380
DB 1321 GTVALSLSYNNMIDNLVENDTWIDVNVVKNITTESDEIOKGELIENILSKLINEDN 1380
QY 1381 KIILNHTNPFYGDINESNRFISLTFSLIEDNIIIEIDLVSYSKYLISGCMKLIENS 1440
DB 1381 KIILNHTNPFYGDINESNRFISLTFSLIEDNIIIEIDLVSYSKYLISGCMKLIENS 1440
QY 1441 SDTQOKIDHIGNGEHQKVIPIYSYDNTKNGFYDYSKKEGLFTAAPSNEIIRNIYFP 1500
DB 1441 NHIQKIDYIGNSLSEKQKIPYSFVDSKNGFNGSTKEGLFVSELFDVVLISKVYMD 1500
QY 1501 DSNLLIYSSKOLADRIINKGDKVLLIGNYKPKDMKVSLSFTIEDNTIKLNGVYLDN 1560
DB 1501 DSNLLIYSSKOLADRIINKGDKVLLIGNYKPKDMKVSLSFTIEDNTIKLNGVYLDN 1560
QY 1561 GVAQIILKFNWAKSALNTSNLMNPLESINIKNIFVNNLDNPIEFITDNTFIISGNSIG 1620
DB 1561 GVAQIILKFNWAKSALNTSNLMNPLESINIKNIFVNNLDNPIEFITDNTFIISGNSIG 1620
QY 1621 QFELICDKNIOPIYFINKIKETSITLYVGNRQNLIVPEPSYHLDSDGNSISSTVINFSOK 1680
DB 1621 QFELICDKNIOPIYFINKIKETSITLYVGNRQNLIVPEPSYHLDSDGNSISSTVINFSOK 1680
QY 1681 YLYGIDRYKVKVLIAPNLYT 1700
DB 1681 YLYGIDRYKVKVLIAPNLYT 1700
QY 1699 YLYGIDSCVKNKVIISPNYT 1699
DB 1699 YLYGIDSCVKNKVIISPNYT 1699

```

```

RESULT 2
US-08-405-496A-10
; Sequence 10, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUTROXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CABROLL, LLP
; STREET: 230 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO

```

STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-10

Query Match 77.9%; Score 6759.5; DB 2; Length 2366;
Best Local Similarity 76.6%; Pred. No. 0;
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;
Qy 1 MNLVNAQLQRMVYKFRIDEVVAITLNALEEHNNMSSVVEKYLKADINLNDYIL 60
Db 1 MSLVNRQLEKMANVRFTQDEYVAILDALEEHNNMSENTVVEKYLKADINSLDIYI 60
Qy 61 NTYKSGRNKALKKFKKYLTVLEVLKXNSLTPVEKNLHFIWGGQINDTAINYINQWKD 120
Db 61 DTYKSGRNKALKKFKKYLTVLEVLKXNNLTPEKNLHFVWIGQINDTAINYINQWKD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLTKTIVSATNTLTSPRENLNDPEPDYKFKRMEIY 180
Db 121 VNSDYNVNVFDSNAFLINTLTKTVVESAINDLTSPRENLNDPRPDYKFKRMEIY 180
Qy 181 DKQKHFDIYKQSEENPEFTIDNIKTYSNEYSKDLEALNKYIEESLNKITTANNNDI 240
Db 181 DKQNFYNYKAQSEENPELLIIDIIVKTYLSNEYSKIDEINLYIEESLNKIQNSGNDV 240
Qy 241 RNLEKFADEDLVRLYNGELVERWNLAAASDILRISMLKEDGGVYLDVVDILPGIQPLFKS 300
Db 241 RNPEEFNGESFNLYEQELVERWNLAAASDILRISALKEIGMYLDVMDLPGIQPLFES 300
Qy 301 INKPDSTINTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEVOESPESALSSKDKSEIF 360
Db 301 IEPSSVTVDFWMTKLEAIMKYKEYPEYTSFMDLDEVOSSPESVLASXSKSEIF 360
Qy 361 LPLDDIKVSPLEVKIAFANNVINOALISLKDSCSLVINQIKRKYKILNDNLNPSINE 420
Db 361 SSLGDMASPLEVKIAFNKGIINQGLISVKDSCSLNIVKQENRYKILNNSLNPAISE 420
Qy 421 GTDFNTMTKIFSDKLASINEDNNMFMKITNYLKVGFPDVRSTINLSGFGVYTGAYQD 480
Db 421

Db 421 DNDFTNTTTFIDSIABANADNGRFMVELGKYLRVGFFDFDKTTINLSGPEAYAAAYQD 480
Qy 481 LLMPKDNSTIHLLEPELRNFEFPKTKISQTEQEIITSLMSFNQARAKSFEEYKKGYFE 540
Db 481 LLMPKEGSMNHLLEADLRNFEISKTNISOSTEGEMASLWSDARAKAQFEYKRYFE 540
Qy 541 GALGEDNDLPAQNTVLDKQYVSKILLSSMKTRNKEIHHVIVQLOQDKISYEASCNLPSK 600
Db 541 GSLGEDNDLPSQNVVDKEYLLEKISSLARSSRGYIHYIVQLOQDKISYEACNLFK 600
Qy 601 DPYSILYOKNIEGSETAYYYYVADAEIKEDKRIIPVOISNKENIKLTIHGHKSRENT 660
Db 601 TPDYSLVFKNIEDSEIAYYNPBGGEIQEIDKTKIPISIDRPRKILKTFIGHGKDFNT 660
Qy 661 DTEANLDVSLSSSEIETILNADKADISPKYIETINLGNMFSYSISAEETYPGKLLKIK 720
Db 661 DIFAGFDVDSLSTEIEAAIDAKEDISPKSIEINLGNMFSYSINVEEYYPGKLLKVK 720
Qy 721 DRVSELMPISIQDSITTVSANQYEVRIINEEGKREILDHSGKWINKKEESIKDISKEYISF 780
Db 721 DKISELMPISIQDSIIVSANQYEVRIINSEGRRELLDHSGEWINKKEESIKDISKEYISF 780
Qy 781 NPKENKIIIVKSKYLHELSTLLOETRNANSSDIDLEKKVMLTECEINVASNIDQIVEGR 840
Db 781 NPKENKIVKSKNLPSTLLOEIRNNSNSSDIDLEEKVMLTECEINVISNIDQIVEER 840
Qy 841 IEEAKNLTSDSINVKIEFKLIESISDLYDLKHONGLDDSHFISFEFISKTEGFRIRF 900
Db 841 IEEAKNLTSDSINVKIEFKLIESISDALCOLKQONELEDSHFISFEFISDSDGFSIRF 900
Qy 901 INKETGNSIFETEKEIPSEYATHISKESINIKOTIFDNVNGKLVKKNLDAAEVNTLN 960
Db 901 INKETGESIFVETETKTFSEYANHTBEISKIKGTIFDTVNGKLVKKNLDTHEVNTLN 960
Qy 961 SAFFIQSLIEYNTTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 AAFPIQSLIEYNSKESLSNLVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDETID 1020
Qy 1021 LLPTLSEGLPIIATIDGVSLGAAIKELSETNDPILLRQEIETAKGIMAVNLTAATASTVT 1080
Db 1021 LLPTLSEGLPIIATIDGVSLGAAIKELSETSDPILLRQEIETAKGIMAVNLTAATAIT 1080
Qy 1081 SALGATASFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETGAFTLLD 1140
Db 1081 SSLGATASFSILLVPLAGISAGIPSLVNNELVLDRDKATKVVVDYFKHLSVTEGVTLLD 1140
Qy 1141 DKIIIMPQDLDVLSEIDFNNSITLCKCEIMRAEGSGHLLTDDIDHFFSPSITIRKPWL 1200
Db 1141 DKIIIMPQDLDVLSEIDFNNSIVLCKCEIMRWEGSGHVTDDIDHFFSAPSIIVREHL 1200
Qy 1201 SIYDVLNLIKKEKIDFSKOLMVLNAPNRFVGYEMGWTFGRSLNDGFKLLDRIDHYEG 1260
Db 1201 SIYDVLVQKEELDLSKOLMVLNAPNRFVFAWETGWTGLRSLNDGFKLLDRIDNYEG 1260
Qy 1261 QYTWYFAFIADALITKLPRIEDTNVINLDGWTSPFIVVITTEQTKNLSVSYFGSG 1320
Db 1261 EFWYFAFIADALITTKPRIEDTNIRINLDSNTRSPFIVPITTEYREKLSVSYFGSG 1320
Qy 1321 GSYLSLSPYNNIDNLAVENDTWIVDNNVNNITTESDEIOKGLTENILSKLINBDN 1380
Db 1321 GYALUSLQYNGNINLIESEDVMDIIVDNNVRVDTIESDKIKKGLLEGILSTLISEN 1380
Qy 1381 KIILNNHINIFYGDINESNRFISLTFSILEDINIIIEIDLVSYSKYILLSGNCKMLIENS 1440
Db 1381 KIILNHNINSEGEVNGSGFVSLTFSILEGINAIEVDLLSKSYKLLISGELKILMNS 1440
Qy 1441 SDIQQKHIGHGNBHQKVIPIYSYIDNETKNGFIDYSKKGGLFTAEPNSIESIIRNIYMP 1500
Db 1441 NHIQKIDYIGFNSBLQKNIPYFVDSEKENGFGINGSTKGLFVSELPDVVYLIKSVTMD 1500
Qy 1501 DSNNLFIYSKDKLDIRINKGDVKLLIGNYFKDDMKVLSLFTIEDTITKINGVYLDEN 1560
Db 1501 DSKPSFGYYSNNLKDVKVITKDNVNILTYGVLKDDIKISLSLTQDEKTIKILNSVHDES 1560

MOLECULE TYPE: protein
US-08-915-136-10
Query Match 77.9%; Score 6759.5; DB 3; Length 2366;
Best Local Similarity 76.6%; Pred. No. 0;
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;
QY 1 MNLVKAQLOKVVYKPKIQEDEVVAIALNALLEEYHNMSESSVVEKYKLDKIDNNLTNYL 60
Db 1 MSLVNRKQLEKXNVFRFQEDYVAIILDALBEYHNMSENVTVEKYKLDKIDNNLTNYL 60
QY 61 NTYKSGRNKALKPKKEYLTMEVLEKNSLTPEVKNLHFIWIGQINDTAINYNQWD 120
Db 61 DTYKSGRNKALKPKKEYLTMEVLEKNSLTPEVKNLHFIWIGQINDTAINYNQWD 120
QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLTLESFRENLDNPFDPKVKRKEMLIY 180
Db 121 VNSDYNVNVFYDSNAFLINTLTKTIVESAINDLTLESFRENLDNPFDPKVKRKEMLIY 180
QY 181 DKQKHFIDYKSOIENPEFIIDNIKTVLSNEYSKOLEALNKYIEESINKITANGNDI 240
Db 181 DKQKHFIDYKSOIENPEFIIDNIKTVLSNEYSKOLEALNKYIEESINKITANGNDI 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVDPILQIGDLPKS 300
Db 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVDPILQIGDLPKS 300
QY 301 INKPSITNTSWEMIKLEAIMKYIYIPGYSKNTFMDLDEEVQSFESALSXSXSSEIF 360
Db 301 INKPSITNTSWEMIKLEAIMKYIYIPGYSKNTFMDLDEEVQSFESALSXSXSSEIF 360
QY 361 LPDDIKVSPLEVKAFANNVINOALSLKDSYCSDLVINOIKRYKILNLDNLPISNE 420
Db 361 LPDDIKVSPLEVKAFANNVINOALSLKDSYCSDLVINOIKRYKILNLDNLPISNE 420
QY 421 GTDNTMTKIFSDKLASISNEDNMFMKITTNYLVKGFAPDVVRSTINLSGGVYTCAYOD 480
Db 421 GTDNTMTKIFSDKLASISNEDNMFMKITTNYLVKGFAPDVVRSTINLSGGVYTCAYOD 480
QY 481 LLMFKDSTNIHLLEPELRNPFPTKISQLEQBITSLWSNQARAKSQPFEYKKGYPF 540
Db 481 LLMFKDSTNIHLLEPELRNPFPTKISQLEQBITSLWSNQARAKSQPFEYKKGYPF 540
QY 541 GALGEDNDLDFRAQNTVLDKDYVSKILLSMKTPNKEYIHYIYVLOQDKIYSKASCNLF 600
Db 541 GALGEDNDLDFRAQNTVLDKDYVSKILLSMKTPNKEYIHYIYVLOQDKIYSKASCNLF 600
QY 601 DYSSTLYOKNIEGSETAYVYVADAEIKEDKIRIPYOISNKRNIKLTFIGHKGSEFNT 660
Db 601 DYSSTLYOKNIEGSETAYVYVADAEIKEDKIRIPYOISNKRNIKLTFIGHKGSEFNT 660
QY 661 DTFANLDVDSLSSEIETILNADKADISPKYIEINLLGCMFYSYSAAETYPKLLKTK 720
Db 661 DTFANLDVDSLSSEIETILNADKADISPKYIEINLLGCMFYSYSAAETYPKLLKTK 720
QY 721 DRYSELMPISQDSITVSANQYEVNRNEEGKREILDHSGKWNKESIKDKSSKEYISF 780
Db 721 DRYSELMPISQDSITVSANQYEVNRNEEGKREILDHSGKWNKESIKDKSSKEYISF 780
QY 781 NPENKIIVKSKYLHELSTLLOQRNANSSDIDLEKKWLTCEINVASNTDROIVEGR 840
Db 781 NPENKIIVKSKYLHELSTLLOQRNANSSDIDLEKKWLTCEINVASNTDROIVEGR 840
QY 841 IEAKNLTSDSINVYKNEFKLIESISDLSYDLKHQGLDDSHFISFEDISKTEGFRIF 900
Db 841 IEAKNLTSDSINVYKNEFKLIESISDLSYDLKHQGLDDSHFISFEDISKTEGFRIF 900
QY 901 INKETGNSIFETEKEIFSEYATHISKEISNTKDTIPDNVNGKLVKKVNLDAAEVNTLN 960
Db 901 INKETGNSIFETEKEIFSEYATHISKEISNTKDTIPDNVNGKLVKKVNLDAAEVNTLN 960
QY 961 SAPPFQSLIEYNTTKESLSNLSVAMKVQVYAOIPSTGLNITITASKVVELVSLDETID 1020
Db 961 SAPPFQSLIEYNTTKESLSNLSVAMKVQVYAOIPSTGLNITITASKVVELVSLDETID 1020

QY 1561 GVAQILKPMNAKALNATNSLMFLSINIKNFYNNLDNPNTEFIIDTNYLFGSNGSIG 1620
Db 1561 GVAQILKPMNR-KGNTNTSDSLSMFLSNGNKSIFVNFLOSNIKFIIDANFIISGTSIG 1619
QY 1621 OFELICCDKXNIQPIYFNFKIKETSYLYVGNRQNLIVPEFSYHLLDSCNISTVINFSQK 1680
Db 1620 QFEFICCDENDIQPIYFNFKIKETSYLYVGNRQNLIVPEFSYHLLDSCNISTVINFSQK 1679
QY 1681 YLYGIDRYNVKVIAPNLYT 1700
Db 1680 YLYGIDSCNVKVISPNYI 1699
RESULT 3
US-08-915-136-10
Sequence 10, Application US/08915136
Patent No. 6290960
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESS: MELEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear

961	Db	AA	PI	Q	L	I	B	N	S	K	S	L	N	S	V	A	M	K	V	Q	V	A	Q	F	S	T	G	L	N	T	I	D	A	A	K	V	V	E	L	S	T	A	L	D	E	T	I	D	1020								
1021	Qy	LL	PT	L	S	E	G	L	P	I	A	T	I	D	G	V	S	L	G	A	A	I	K	E	S	E	T	N	D	P	L	L	R	Q	E	I	E	A	K	I	G	M	A	V	N	L	T	A	A	S	T	A	I	V	T	1080	
1021	Db	LL	PT	L	S	E	G	L	P	I	A	T	I	D	G	V	S	L	G	A	A	I	K	E	S	E	T	N	D	P	L	L	R	Q	E	I	E	A	K	I	G	M	A	V	N	L	T	A	T	A	I	T	1080				
1081	Qy	SA	L	G	I	A	S	G	S	I	L	V	P	L	A	G	I	S	A	G	I	P	S	V	N	N	E	L	I	I	O	D	K	A	T	V	I	D	P	F	K	H	I	S	L	A	E	T	S	G	A	T	L	L	D	1140	
1081	Db	SS	L	G	I	A	S	G	S	I	L	V	P	L	A	G	I	S	A	G	I	P	S	V	N	N	E	L	I	V	L	R	D	K	A	T	V	I	D	P	F	K	H	I	S	L	A	E	T	S	G	V	T	L	L	D	1140
1141	Qy	DK	I	M	P	O	D	D	L	V	E	I	D	F	N	N	S	I	T	L	G	K	E	I	W	R	A	G	S	G	S	G	H	T	I	D	D	I	D	H	F	F	S	P	S	I	T	Y	K	P	W	L	1200				
1141	Db	DK	I	M	P	O	D	D	L	V	E	I	D	F	N	N	S	I	V	L	G	K	E	I	W	R	M	G	G	S	G	H	T	V	T	D	I	D	H	F	F	S	A	P	S	I	T	Y	R	E	P	H	L	1200			
1201	Qy	SI	Y	D	V	L	N	I	K	E	K	I	D	F	S	K	D	L	M	V	L	N	A	P	N	R	V	P	G	Y	E	M	G	T	P	F	R	S	L	D	N	O	G	T	K	L	L	D	R	I	O	H	Y	E	G	1260	
1201	Db	SI	Y	D	V	L	E	V	O	K	E	I	D	L	S	K	D	L	M	V	L	N	A	P	N	R	V	F	A	W	E	T	G	T	P	G	L	R	S	L	E	N	D	G	T	K	L	L	D	R	I	O	N	Y	E	G	1260
1261	Qy	Q	F	Y	R	Y	P	A	F	I	A	D	A	L	I	T	L	K	P	R	Y	E	D	T	N	R	I	N	L	D	G	N	T	R	S	F	I	V	P	T	T	E	Q	I	R	K	N	L	S	I	S	F	Y	G	S	G	1320
1261	Db	E	F	Y	R	Y	P	A	F	I	A	D	A	L	I	T	L	K	P	R	Y	E	D	T	N	R	I	N	L	D	S	N	T	R	S	F	I	V	P	T	T	E	R	E	K	L	S	I	S	F	Y	G	S	G	1320		
1321	Qy	G	S	Y	S	L	S	P	Y	M	M	I	D	L	N	V	E	N	D	T	W	I	D	V	N	N	V	K	N	I	T	E	S	D	E	O	K	G	E	L	I	E	N	I	L	S	K	L	M	I	E	D	N	1380			
1321	Db	G	T	Y	A	L	S	Q	Y	N	G	I	N	E	S	D	V	M	I	D	V	N	V	R	D	V	I	E	S	D	K	K	G	D	I	E	G	I	L	T	S	E	E	N	1380												
1381	Qy	K	I	L	N	N	H	T	I	N	F	G	D	I	N	E	S	F	I	S	L	T	S	I	E	L	D	I	N	I	I	E	I	D	L	V	S	K	Y	K	I	L	L	S	G	N	C	K	L	I	E	N	S	1440			
1381																																																									

RESULT 4

US-08-957-310-10
Sequence 10, Application US/08957310
Patent No. 6365158
GENERAL INFORMATION:
APPLICANT: Williams, James A.
APPLICANT: Kink, John A.
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
TITLE OF INVENTION: DISEASE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

541	Db	GSIGBDDNLDFSQNI	VVDKEYLLEKISSI	ARSERGVIIHYI	VQLOQDKISYEAA	CNLF	PAK	600					
601	Qy	DPYSSILYQKNIEG	SETAYYYVADAEI	KEDIKYRIPYQIS	NRKNIKLTFI	GHGSEFNT	660						
601	Db	TPYDSVLFOKNIE	DSIAYYNPGDGEI	QEDIKYKIPSII	SDRPKIKLTFI	GHGKDEFT	660						
661	Qy	DTFANLDVDSLS	EIEITINLAKADIS	PKYIHNLLGCNM	FYSYSABE	TYPKLLLIK	720						
661	Db	DIPAGFDVDSL	TEIEAAIDLAKED	ISPKSIEINLLGC	NFYSINVEE	TYPKLLLVK	720						
721	Qy	DRYSELMPSIT	SODSITVSANQY	VRINERGBEIL	DHSGWKINKE	BSIIKDISSEK	YSIF	780					
721	Db	DKISELMPSIS	QDSIIIVSANQY	VRINSGRELL	DHSGEWINK	BEESIIKDISSEK	YSIF	780					
781	Qy	NPXENKIIIVKS	XYLHBLSTLQBI	RNNANSSDIDLE	KVMLTECEIN	VASNDRQI	VEGR	840					
781	Db	NPXENKITVKS	XNLPSTLQBI	RNNANSSDIELE	KVMLTECEIN	VINISNIDQI	VEER	840					
841	Qy	IEBAKNLTS	DSINYIKNEFKL	IESISDSLYDL	KHONGLODS	HPISPEDISK	TENGFRIR	900					
841	Db	IEBAKNLTS	DSINYIKDEFKL	IESISDALCDL	KQNELED	SHFISPEDI	SETDEGFS	IRF	900				
901	Qy	INKETGNSIF	ETETEKIFSEY	AHISKEISNI	KDITIFDNV	GKLVKVNLA	DAHEVNTLN	960					
901	Db	INKETGESIF	ETETEKIFSEY	AHITTEESIK	IGTIFDVTNG	KLVKVNLD	THEVNTLN	960					
961	Qy	SAFFIOSLIE	YNTTKESLS	NLVAMKVQVYA	QLFSTGLNT	ITDASKV	ELVSTALDE	ITD	1020				
961	Db	AAFFIOSLIE	YNSKESLS	NLVAMKVQVYA	QLFSTGLNT	ITDAAKV	ELVSTALDE	ITD	1020				
1021	Qy	LLPTLSEGL	PIIATIIDGVS	LGAAKELST	NDPLLROBIE	AKIGIMAVNL	TAASTAIVT	1080					
1021	Db	LLPTLSEGL	PIIATIIDGVS	LGAAKELST	SDPLLROBIE	AKIGIMAVNL	TAATTAIT	1080					
1081	Qy	SALGIASG	FSILLVPLAGI	SAGIPSLVNN	ELLIQDKAT	KVIDYFKHIS	LAETEGA	FTLTD	1140				
1081	Db	SSIGIASG	FSILLVPLAGI	SAGIPSLVNN	ELVLRDKAT	KVDYFKHVS	IVETEGV	FTLTD	1140				
1141	Qy	DKIIMPODD	LVLSIEDFN	NNSITLGKCEI	WRABEGSG	HTLTD	DIHFFSP	SPTITTKPWL	1200				
1141	Db	DKIIMPODD	LVLSIEDFN	NNSIVLGKCEI	WRABEGSG	HTVTD	DIHFFSA	PSITTYREPHL	1200				
1201	Qy	SIYDVLN	IKKEKIDF	SXOLMVLN	APNRVFGY	EMGWTGPR	SRLNDGT	KLDIRPHYSG	1260				
1201	Db	SIYDVLE	VQKEELD	LSXOLMVLN	APNRVFAW	TGTPGLSR	LENDGT	KLDIRONYEG	1260				
1261	Qy	QFYWRYE	AFPTADALIT	LKPRYEDTV	NRINLDGNT	SRFSFIVPV	LITTTQIR	KNLISYSYGG	1320				
1261	Db	EFWRYE	AFPTADALIT	LKPRYEDTV	NRINLDN	STRFSFIVP	ITITTEYI	REKLSYSYGG	1320				
1321	Qy	GSYSLS	SPYNNIDL	NVENDTWI	DVDNVKN	ITTESDIE	IOKGLIEN	ILSKLI	EDN	1380			
1321	Db	GTVALS	QVNMGINIE	LSGDWII	DVDNVRD	VTTIESDKI	KKGDLIE	GLSTL	GIEEN	1380			
1381	Qy	KIILNHT	INIFYGDINES	NRFISLTF	SILEDINIIE	IBLVKS	YKILLSGN	CMKLIENS	1440				
1381	Db	KIILNSH	INIFSGEVNG	SFVSLTFP	SILEGINAI	IEVDLLS	KYKLLIS	GEKLI	MLNS	1440			
1441	Qy	SDTOOK	IDHIGFNGE	HOXYIPYS	VINDETK	NGFDYSK	KEGFTAB	AFBSNESI	IRNIYMP	1500			
1441	Db	NHIOOK	IDYIGFNS	ELOKNI	PIYSPVD	SEKXENG	FINGSTKE	GLFVSEL	PQWILISKYMD	1500			
1501	Qy	DSNNLFY	SSKLDKDIR	INKGDV	KVLLI	GNFKDD	MKVSISFT	IEDNTIK	LVNGVYLDEN	1560			
1501	Db	DSKPSFGY	SNLKDVKVI	KDNVN	ILUTGY	YVKDDIKISL	STLQDE	BKTIK	LVNSVHDES	1560			
1561	Qy	GVAQIL	KFMNNAKS	ALNTS	NLMFLBS	INIKNTFY	NNLD	PNTE	IFILDTNFI	SGNSIG	1620		
1561	Db	GVAIIL	KFMNR	-KGNVT	SDSLMS	FLES	MNAIKSIFV	NFLO	QNIK	FILDA	FIISGTTISIG	1619	
1621	Qy	QFELIC	DKOKNIQ	PIYFN	KITKETS	TYLYVGN	RQNLIV	EPSYH	ILDDSGN	ISSTVIN	FSQK	1680	
1620	Db	QFETIC	BDNIOQ	PIYFN	KITFN	LETNYLY	VGNRQ	MIVPE	NYDILDD	SGD	ISSTVIN	FSQK	1679

Qy 1681 YLYGIDRYNKVILAPNLYT 1700
||||| ||||:||||
Db 1680 YLYGIDSCVNKVISNLYT 1699

RESULT 5
US-10-011-366-10
; Sequence 10, Application US/10011366
; Patent No. 6573003
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT
; OF DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-No. 6573003-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingollia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-011121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-011-366-10

[illegible]

Db 121 VNSDYNVYDSDNAFLNLTAKTVVESAINDTLESFRENLDNPRDYNNKFRKMEIY 180
Qy 181 DKQHFIDYKQIERNPEFIIDNIKTYSNEYSKDLBALNKYIBESLNKITAANGNDI 240
Db 181 DKQKNFYNYKAQREENPELLIIDIVKTVLSNEYSKEIDELNTYIBESLNKIQNSGNDV 240
Qy 241 RNLEKFADEDLVLYNQELVERNLAAASDILIRLSMLKEDGGYDVLDPGIDPLFKS 300
Db 241 RNPEEFKNGESFNLVEQELVERNLAAASDILISALKSIGMYLDDMLPGIDPLFES 300
Qy 301 INKPDSITNTSWEMIKLEAMKYKEVYIPGYTSKNFDMLEEVQORSFESALSSKSKSEIF 360
Db 301 IEPSSVTVDFWMTKLEAMKYKEVYIPGYTSKNFDMLEEVQORSFESALSSKSKSEIF 360
Qy 361 LPDDIKVSLEVKIAPFANNVQALISLKDSCYSDLVINQIKNRYKIILNDLNPISNE 420
Db 361 SSGDMEASPLEVKIAPFANNVQALISLKDSCYSDLVINQIKNRYKIILNDLNPISNE 420
Qy 421 GTPDNTMTKIFSKLASISNEDNMFMKITNYLKYGFAPDVRSTINLSGPGYVYQAYOD 480
Db 421 DNDFTNTTTFIDISMAEADNAGRFMELGKYLVRGVFPDVKTTINTLSGPEAYAAAYQD 480
Qy 481 LLMFKONSTNHLLEPELRNPEPKTKISQITTEBITSLMSFNQARAKSQPEEYKKGYPE 540
Db 481 LLMFKEGSMNHLLEADLRNPEPKTKISQITTEBITSLMSFNQARAKSQPEEYKKGYPE 540
Qy 541 GALGEDDNLDPQONTVLDKDYVSKILLSMKTRNKEYIHYIVOLQDKI-SYASCNLPSK 600
Db 541 GSGEDDNLDPQONTVLDKDYVSKILLSMKTRNKEYIHYIVOLQDKI-SYASCNLPSK 600
Qy 601 DPXSSILYQKNIQEGSETAYYYYVADAEIKEDKIRIPYQISNKNRNIKLTFIGHKSEPT 660
Db 601 TPYDSVLFOKNIQEGSETAYYYYVADAEIKEDKIRIPYQISNKNRNIKLTFIGHKSEPT 660
Qy 661 DTPANLVDLSISBETILMAKADISPKYIEINLLCGNMFYSISABETYPCKLLKIK 720
Db 661 DIFAGFDVLSISBETILMAKADISPKYIEINLLCGNMFYSISABETYPCKLLKIK 720
Qy 721 DRVSELMPSISQDSITVSANOYVRINEEGKREILDSHGWINKESIIKDISSEKVISF 780
Db 721 DKISELMPSISQDSITVSANOYVRINEEGKREILDSHGWINKESIIKDISSEKVISF 780
Qy 781 NPENKIIVKSKYHLSTLLOERNANSSDIDLEKKVMLTECEINVASNIDROQVEGR 840
Db 781 NPENKIIVKSKYHLSTLLOERNANSSDIDLEKKVMLTECEINVASNIDROQVEGR 840
Qy 841 IEEAKNLTSDSINYKNEFKLIESDSLYDLKHQGLDSDHSPISFEDISKTENGPRIRP 900
Db 841 IEEAKNLTSDSINYKNEFKLIESDSLYDLKHQGLDSDHSPISFEDISKTENGPRIRP 900
Qy 901 INKETGNSIETETEKEIFSEYATHISKEISNIDKTFDNDVNGKLVKKNLDAAEVNTLN 960
Db 901 INKETGNSIETETEKEIFSEYATHISKEISNIDKTFDNDVNGKLVKKNLDAAEVNTLN 960
Qy 961 SAPIQSLIEVNTKESLSNLVAMQVVAQFSLGNTITDASKVVELSVTALDETID 1020
Db 961 AAPIQSLIEVNTKESLSNLVAMQVVAQFSLGNTITDASKVVELSVTALDETID 1020
Qy 1021 LLPTLSGLPLIATIIDGVSLGAIKELSETNPLLRQIEAKIGIMAVNMLTAATAIYT 1080
Db 1021 LLPTLSGLPLIATIIDGVSLGAIKELSETNPLLRQIEAKIGIMAVNMLTAATAIYT 1080
Qy 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELIQQDKATKVIDYFKHISLAETGFTLLD 1140
Db 1081 SSGIAGSFSILLVPLAGISAGIPSLVNNELIQQDKATKVIDYFKHISLAETGFTLLD 1140
Qy 1141 DKITMPQDDIVLSIDFNNSITLKGCEIWRAGSGSHTITDDIDHFFSPSITYRKPWL 1200
Db 1141 DKIMPPQDDIVLSIDFNNSITLKGCEIWRAGSGSHTITDDIDHFFSPSITYRKPWL 1200
Qy 1201 SIYDVLNKKKIDFSDKDLVLPNAPNRFVEMGTGTPGRLSDNDGTCLLDRIRDHYEG 1260

Db 1201 SIYDVLVQKEELDLSKDLVLPNAPNRFVEMGTGTPGRLSDNDGTCLLDRIRDHYEG 1260
Qy 1261 QFYRYFAFADALITKLKPRYEDTVRINLDGNTSRFIVPVITTEQIRKNLSYSFYGSG 1320
Db 1261 EFYRYFAFADALITKLKPRYEDTVRINLDGNTSRFIVPVITTEQIRKNLSYSFYGSG 1320
Qy 1321 GSYLSLSPYNNMTDLNLENDTWIDVNVVKNITTESDEIOKGLIENILSKLNIEDN 1380
Db 1321 GTYALSLSQYNGMINIELSDWIIDVNVVRDVTTIESKIKKGDLEGLISTLSIEN 1380
Qy 1381 KIILNHTINPYGFINESNRFISLTSILEDINIIEIDLVSYSKYILLGNCMKLIENS 1440
Db 1381 KIILNSHEINFSGEVNGSGVSLTFSILEGINAIEVDLLSKYKLLISGELKILMNS 1440
Qy 1441 SDIOQKIDHIFNGEHOXKIPYSVIDNETKNGFIDYSKKEGLFTAFESNIESITRNYP 1500
Db 1441 NHIQOKIDYIFGSELQKNIPYSFVDSGEKNGFINGSTKEGLFVSELPVVLISKYVMD 1500
Qy 1501 DSNLFIYSKDLKDIRIINKGDVKLLIGNYFKDDMKVLSFTIEDNTIKLNGVYLDEN 1560
Db 1501 DSKPSFGYSNNLKDVKVITKONVNIUTGYLKKDDIKISLTLQDEKTIKINSVHLOS 1560
Qy 1561 GVAQILKFMNNAKALNTSLSMNFLESINIKNIFNNLDNFIILDTNFIISGNSIG 1620
Db 1561 GVAELKPMNR-KGNTWTSLSMSFLESMAIKSIFVNFLOSNIKIFILCANFIISGTSIG 1619
Qy 1621 QFELICDKXNIQYFNFKIKETSITLYVGNRQNLIVBPSYHLLDGSNISTSTVINFSQK 1680
Db 1620 QFELICENDNIQYFNFKIKETSITLYVGNRQNLIVBPSYHLLDGSNISTSTVINFSQK 1679
Qy 1681 YLYGIDRVNKKVITAPNLYT 1700
Db 1680 YLYGIDSCWKNVISPNIYT 1699

RESULT 6
US-09-084-517-10
; Sequence 10, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992


```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-084-517-10

Query Match 77.9%; Score 6759.5; DB 4; Length 2366;
Best Local Similarity 76.6%; Pred. No. 0;
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 19

QY 1 MNLVNAQOLQKVVYKPIQDEVAIINALAEYVHNMSESSVVEKYLKDKDNNLTNNLT 60
DB 1 MSLVARKQDEKAVNRFQTQDEVAIDALAEYVHNMSENVVEKYLKDKDINSUTDIYI 60

QY 61 NNYKSGRNKALKKFKYELTMEVLKNNSLTPVEKNLHPITWGGQINDTAINYINQWKD 120
DB 61 DRYKSGRNKALKKFKYELTMEVLKNNSLTPVEKNLHPVWIGGQINDTAINYINQWKD 120

QY 121 VNSDYTVKVFYDSNAFLINTLKKTVISATNNTLESFRENLDPEFDYKPKFRKMEIYY 180
DB 121 VNSDYVNVVYDSNAFLINTLKKTVESAINDTLESFRENLDPRFDYKPKFRKMEIYY 180

QY 181 DKQKHFDYKSOIENPEFIIIDNLIITYLSNEYSKDLBALNKYIEESLNKITTANNNDI 240
DB 181 DKQKPFINTYKQARENPELLIIDIVKITYLSNEYSKIDELNYYIEESLNKITTQSGNDV 240

QY 241 RNLEKPADBDLVRNLQBLVERWNLAASAADILRISMLKEDGGVYLDVDILPGIQDLPFS 300
DB 241 RNFEFPKNGESNLVEQLVERWNLAASAADILRISALKEIGGWYLDVDMLPGIQDLPFS 300

QY 301 INKPSITNTSWEMTKLEAIWKYKEYIPGYTSRNFMDLDEEVQSFESALSXSOKSEIF 360
DB 301 IEKPSVTVDFWEMTKLEAIWKYKEYIPEVTSRNFMDLDEEVQSFESVLSKSKSEIF 360

QY 361 LPDLDDIKYSLPVKTAFAANNVINCALISLSDSYCSDLVINQIKRYKILNDNLAFSNE 420
DB 361 SLSGDDMEASPLVKIAFNSKGIINQGLISVSDSYCSNLIVKQIENRYKILNNSLPAISE 420

QY 421 GTDFNTMTKIFSDKLASISNEDNMFMKIKITNLYKGFADVPDASTNLSPGVYTCAYOD 480
DB 421 DNDFTNTTTFDTSIMAZANDNGRPFMELKYLVRGVGFFDVKTITNLSPGPAAYAAQD 480

QY 481 LLMFKNQSTNIHLLEPELRNFPFKTKISQLTQEITSLWSFNQARAKSQFEYKKGYPF 540
DB 481 LLMFKEGSMNHLIEADLRNFEISKTNISQSTQEMASLWSFDDARAKAQFEYKENYFE 540

QY 541 GALGEDNDLDPAQNTVLDKDYVSKKILSSMKTRNKEYIHIYVOLQGDKITSYEASCNLPFSK 600
DB 541 GSLGEDNDLDFSQNTVDPKEYLLEKISLSAESSESGYIHIYVOLQGDKITSYEAACNLFAK 600

QY 601 DRYSSILYKQNIEGSETAYVYVADAEIKEDKYRIPVQISNKRNIKLTFIGHGKEPNT 660
DB 601 TPVDSVLFGQNLDEGEIAYYNGDGEIQEIDKYKIPSIISDPKIKLTFIGHGKDEFTN 660

QY 661 DTFEANDVDSLSSEIETILNKAADISPKYIEINLLGNMFSYSISABETYPGKLLKIK 720
DB 661 DIFAGFDVDSLSTEIEAADIADLAKEDISPKSIEINLLGNMFSYSINVEETYPGKLLKYK 720

QY 721 DRVSELMPISQDSITVSANQYEVTEINBEGKREILDHSGKWINKEESI IKDISSKYIISF 780
DB 721 DKISELMPISQDSITVSANQYEVTEINBEGKREILDHSGKWINKEESI IKDISSKYIISF 780

```

Qy	781	NP	KENKIIVKSYLHELSTLLQEI	RNNANS	SDIDLEKKVWLTECEINVASN	DRQIVEGR	840	
Db	781	NP	KENKITVKSQULPELSTLLQEI	RNNNSNDIELEEKVWLTECEINVSIN	DTQIVEER	840		
Qy	841	IE	EAKNLTSDSINYIKNEFKLIESDSI	YDLKHONG	LDSDSFIPEDISKTEENGRIPE	900		
Db	841	IE	EAKNLTSDSINYIKNEFKLIESD	ALCDLQOONE	LEDSDSFIPEDISEDEGSIRP	900		
Qy	901	IN	KETGNSFIETEKIEFSEYATH	SKELSNIKOTIFDMVNG	KLVKKNLDAAEVNTLN	960		
Db	901	IN	KETGESFVETEKTFSEYANH	ITEELSKIGTIFDVNG	KLVKKNLDTTHEVNTLN	960		
Qy	961	SA	FFIOSLIEYNTTKESLNSLVAM	KVQYQALFSTGLNTITD	ASKVBLVSTALDETID	1020		
Db	961	AA	FFIOSLIEYNSKESLNSLVAM	KVQYQALFSTGLNTITD	AAKVBLVSTALDETID	1020		
Qy	1021	LL	PTLSEGLPIIATITIDGVS	LGAIAKELSETNDPILRQ	IEAKIGHMAVNLTAASAIYT	1080		
Db	1021	LL	PTLSEGLPIIATITIDGVS	LGAIAKELSETNDPILRQ	IEAKIGHMAVNLTAATTAIT	1080		
Qy	1081	SA	GIASGFSILLAVLAGISAGT	PSLVNVELLIQDKATKVIDY	FKHISLAETEGASTLLD	1140		
Db	1081	SS	GIASGFSILLAVLAGISAGT	PSLVNVELVURDKATKV	DYFKVSLVETEGVFTLLD	1140		
Qy	1141	DK	IMPQDOLVISEIDFNNS	SITGLKCEIWRAGEGSGHT	LTDDIDHFFSPSITYRKPL	1200		
Db	1141	DK	IMPQDOLVISEIDFNNS	SITGLKCEIWRAGEGSGHT	VTDDIDHFFSAPSITYREPH	1200		
Qy	1201	SI	YDVNLINKEKIDFSKDL	MVLPNAPRVFGEMGTP	FRSLDNDGTLLDIRDHYEG	1260		
Db	1201	SI	YDVLEVOKEELD	SKDLMLPNAPRVFAWET	GTGLRSLENDGTLLDIRDNYEG	1260		
Qy	1261	QF	WRYPAFIADALITLKP	RYEDTVNRILNDGNTES	FIVPVITTEQIRKNLSYSVSG	1320		
Db	1261	EF	WRYPAFIADALITLKP	RYEDTVNIRLDSNTRSF	IVPIITTEYIREKLSYSPYSG	1320		
Qy	1321	GS	YSLSPANNMIDLXVEND	TWVIDVNVKNIETES	DEIQKGLIENILSKLNIEDN	1380		
Db	1321	GT	YALSQYNNGINIELSE	SDVWIIDVNVVRDVIT	IESDKKGGDIEGILSTLSEEN	1380		
Qy	1381	KI	LNHTINFYGDINESR	FISLTESI	EDINIIIEIDIVSKSYKILLSGNCKMLIENS	1440		
Db	1381	KI	LNSHIEINFGSEVNG	SGFVSLTFSILEGINA	IEVDLLSKSYKILLSGELKILMLNS	1440		
Qy	1441	SD	IQKIDHIGFNGEHQYIP	YSYIDNETKYN	GFIIDYSKKEGLFTAFBNESIIIRNIYMP	1500		
Db	1441	NH	IQKIDYIGFNSSELQNI	PYSFVDSGKENG	FINGSTKEGLFVSELDPDWLISKVYMD	1500		
Qy	1501	DS	NNLFYYSKDLKDR	IRINKGVDKLLIGNY	FKDDMKVLSFTIEDTNTIKLVGYLDEN	1560		
Db	1501	DS	XPFGYQNNLKVDKVIT	KONVNILGTGYLXDD	IKISLSLTQDEKTIKLSNVHLDES	1560		
Qy	1561	GV	AILKFMNKA	SALNTSLSNLNFELES	INIKNI FYNNLDPNT	EFILDTNFIISGNSIG	1620	
Db	1561	GV	AILKFMNR	KONTWTSOSLSFLES	NMKSIFVNFLOSNI	KFILDANFIISGTSIG	1619	
Qy	1621	QF	ELICDKXNIQPYFIN	FKETSYTYLVGN	RQNTLIVEPSYHLD	DSGNTSSVTINFSQK	1680	
Db	1620	QF	BEICDENNIQPYFI	KFTLTETNYTYLVGN	QNMIVERNYD	LDDSGDISSTVINFSQK	1679	
Qy	1681	YL	XGIDRYNVKVI	APNLNYT	1700			
Db	1680	YL	XGIDSCVKNWIS	PNIYT	1699			

RESULT 7

US-08-480-604A-6

; Sequence 6, Application us/08480604A

; Patent No. 5736139

; GENERAL INFORMATION:

; APPLICANT: KINK, JOHN A.

; APPLICANT: THALLEY, BRUCE S.

RESULT 7
US-08-480-804A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.

APPLICANT: PADHYE, NISHA V.
 APPLICANT: FIRCA, JOSEPH R.
 APPLICANT: STAFFORD, DOUGLAS C.
 TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
 PREVENTION OF C. DIFFICILE DISEASE
 NUMBER OF SEQUENCES: 32
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MEDLEN & CARROLL, LLP
 STREET: 220 MONTGOMERY STREET, SUITE 2200
 CITY: SAN FRANCISCO
 STATE: CALIFORNIA
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/480,604A
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/422,711
 FILING DATE: 14-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/405,496
 FILING DATE: 16-MAR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/329,154
 FILING DATE: 25-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/161,907
 FILING DATE: 02-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/985,321
 FILING DATE: 04-DEC-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/429,791
 FILING DATE: 31-OCT-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: INGOLIA, DIANE E.
 REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPHD-01763
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-480-604A-6
 Query Match 51.58; Score 4467; DB 1; Length 2710;
 Best Local Similarity 50.78; Pred. No. 2.2e-262;
 Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;
 QY 1 MNLVNAQLQKVVYKFRIOEDYVAINALAEYHNHNSSESVYKYLKLDINNLDNYL 60
 DB 1 MSLISKELIKLAY-SIRPRENEVYKTLTNLDYKNTNNENKYLQKLKINESIDVFM 59
 QY 61 NTKYKSGRNKALKKFKYLVMEVLELKNLSLTPVERKNLHFVITGGQINDTAINYNQWKD 120
 DB 60 NKYTKSGRNKALSLLKDLKLVILKNSNTSPVERKNLHFVITGGQINDTAINYNQWKD 119
 QY 121 VNSDYTKVFDVSNAPLNTLTKKTVESATNTLESFRENLDNPEEDYKFKRWEIY 190
 DB 120 INAEYNKLMWDSFAVLNLLKKAIVSSSTTEALQLLEESIQNPQFNMFKFKRWEIY 179
 QY 181 DKQKHFDYKQSEENPEFTIDNIITKYLNSNEYSKDLKALNKYIESLNKITTANNNDI 240

180 DRQKRFINYKQKINPQVPTIDDIKSHLVSEYNRDETVEISYRTNSLRKINSNHDGI 239
 QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRLSMKEDGGVYLDVLDLPGIOPDLFKS 300
 DB 240 RANSLFTEQELLNIYSQELNRLNGLAAASDIVLLALKNFGVYLDVLDVLMPLGHSDFKT 299
 QY 301 INKPDSTINTSWEMIKLEAIKMYKEYIPGYTSKNFDMDEEVQSPESALSSKSDKSEIF 360
 DB 300 ISRPSISGLDRWEMIKLEAIKMYKYNNTYTSNFKLDQOLKDNFKLIIESKSEIF 359
 QY 361 LPLDDIKVSPLEVKIAFANNVINQALISLKDSCYSDLVINQIKRYKILNDNINPSINE 420
 DB 360 SKLENLVSDLEIKIAFALGVLINQALISKQSGVLTNLVIEQVKRYOFLNQHLPAS 419
 QY 421 GTDENTMFKPSDKLASISNEDNMFMKITYNLKVGFPADPVRSTINLSGPGVYTGAYD 480
 DB 420 DNNFTDTTKIFHDSFNLSATAENSMTLKIAPLQVGFPEARSTLSLQSGFAYASAYD 479
 QY 481 LLMEKDNSTNHLLEPELRNPEPKTKISQITQETLSLWSEFNQARAKSOFEEYKKGFFE 540
 DB 480 FINQENTIEKTLKASDLIEFKPENLSQITQETLSLWSEFNQARAKSOFEEYKKGFFE 539
 QY 541 GALGEDDNLDPQANTVLDKDY-VSKILSS--MKTRNKEYIHYIVQLQGDKISYEASCNL 597
 DB 540 GSLSENGVDENKNTALDKYLLNNKIPSNVVEAGSKNYVHYIQLQGDKISYEACNL 599
 QY 598 FSKDPYSIIYQKNIQEGSEYAYYYVAD--AEIKEDKYPQVQISNKRNIKLTFFIGHGK 655
 DB 600 FSKNPKNSIIQRNW--NESAKSYFLSDGSEILELNKRYRIPERLKNKKEKVKVTFIGHGK 657
 QY 656 SEFNTDTFANLDVDSLSESEITINLAKADISPKYIEINLLGCNMPYSISAEETYPGKL 715
 DB 658 DEFNTSEFARLSVDSLSEISSEFLDTIKDISPKNVEVLLGCNMPYSIDFNVETYPGKL 717
 QY 716 LLKIKDRVSELMPSISODSITVSANOYEVINEGKREILDHSGKWNKKEESIKDSSK 775
 DB 718 LLSIMDKITSLPDVKNKNSITIGANQEVINESEKRELLAHSGKWNKKEEAINMSDSSK 777
 QY 776 EYISFNPKENKIIIVKSKYLHELSTLLOEIRNANSSDIDLEKKVMLTECEINVASNIDRQ 835
 DB 778 EYIFDSIDNKLKAKSKNIPGLASISEDITLLDASVSPPTKFLANLKNLNISSIGDY 837
 QY 836 IVEGRIBKAKNLTSDSINYKNPEKLESISDSIYDLKHQGLDSDHIFIPEDISKETENG 895
 DB 838 IYVEKLEPVKNIHNSIDDLIDFNLENVSDLEYELKLNLDKYLISPEDISKONST 897
 QY 896 FRIRPKNETGNSIFIEKEIFSEYATHISKETSNIKDTIFDNVNGKLVKKVNLDAHE 955
 DB 898 YSVRFINKSGESVYVETEKEIFSKYSEHTKEISTIKNSIITDVNGLDNIQDHTSQ 957
 QY 956 VNTLNSAFFIOSLJEYNTTKESLNSLVAMKVQVYQALFSTGLNTITDASKVVELVSTAL 1015
 DB 958 VNTLNAFFIOSLIDYSSNKVDLNDLSTSVKQVYQALFSTGLNTIYDSIQLVNLSNAV 1017
 QY 1016 DETDLLPTLSEGLPIIATIIDVSGLGAIXELSETNDPLLRQIBAKIGIMAVNLTAAS 1075
 DB 1018 NDTINVLITTEGPIVSTILDINGLGAIXELDEHDPDLKKELEAKVGLAINMSLSI 1077
 QY 1076 TAVTSALGASGSPILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETGA 1135
 DB 1078 AATVASIVGIGAEVTFILPITAGISAGIPSLVNNELILHDKATSVMYVNFLESKKYGP 1137
 QY 1136 FTLLDDKILMPQDDLVLSEIDFNNSITLKGCEIWRABGGSGHTLTDDIDHFFSPSPITY 1195
 DB 1138 LKTEDDKILVPDLDLVISEIDFNNSIKLGTGTCNLAIEGGSGHVTGNIDHFFSPSPISS 1197
 QY 1196 RKPWLSIYDVNLINKEKIDFSKDLMLVLPNAPNRFVGYEMGWTGPRSLDNDGTLLDRIR 1255
 DB 1198 HPSLSIYSAGIETENLDFSKIMLNPANRPFVFWETGAVPGLRSENGDTRLLDSIR 1257
 QY 1256 DHYEQFYWRYPAFIADALITKLKRYEDTNRINLDGNTKRSFIVPVITTOIRKNLSYS 1315
 DB 1258 DLYPGKFWRFYAF--DYAITLKVYEDTNIKIKDKDTRNFIMPTITTTNEIRKNLSYS 1316

[illegible]

RESULT 9

```

US-08-915-136-6
; Sequence 6, Application US/08915136
; Patent No. 6230960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FRICA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND
; TITLE OF INVENTION: PREVENTION
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL,
; STREET: 220 MONTGOMERY STREET,

```

QY	361	LPDDDIKVSPEVKIAFANNKSVINQALISKDSYCSDLVINOIKRYKILNDNLNPSINE	420
DB	360	SKLENLVSDLEIKIAFALGVSINQALISKQSYLTNLVIEQVKRYQFLNQHLPATES	419
QY	421	GTDPTNTMKIIPSKKLASIGNEDNMFWIKITNLKYGFAPDVRSTINLSGPGVYTGAYD	480
DB	420	DNNTDITTKIFHDSLFNSATAENSMTLTKIAPYLQGFWPEARSTISLSPGAYASAYD	479
QY	481	LLMPKONSNIHLEPELNFPEPKYKISOLTEQBITSLWSFNQABAKSQPEYKKGYPE	540
DB	480	FINLOQNTIEKTLKASDLTEFKPEPNLNSOLTEQINSLWSFDQASAKYQFKYVRDYG	539
QY	541	GALGEDNDLDFQONTVLDKDY-VSKKILSS--MKTRNKEYIHYIVOLQGDKISYEASCNL	597
DB	540	GSISENGVDPNKNTALDKNYLLNKKIPSNVNEZAGSKYVHIILOQDDISYEATCNL	599
QY	598	FSKDPYSSILYQKNIEGSETPAYYYYVAD--AEIKEIDKYRIPYQISXKNRIKLTFTGHOK	655
DB	600	FSKNPKNSIIIQRM--NESAKSYFLSDGCEISLELNKYRIPERLKNKEKVKVTFIGHOK	657
QY	656	SEFNTDTPANLVDVSLISSEIETILNLAKADISPKYIENILNGNMESYSISAEETYPGKL	715
DB	658	DEFNTSEFARLSVDSLSNEISSLDTIKDISPNVEVNLGNCMFSDYFNVEETYPGKL	717
QY	716	LLKIKORVSELMPSISODSTIVSANQYEVRAINEEGKREILDHGKWKINEESIIDKISSK	775
DB	718	LLSIMGKITSLPDVKNKNSITIGANQYEVRAINSEGRKELLAHGKWKINEEAIMSDLSK	777
QY	776	EYISFNPKENKIIVKSKYLHELSTLLOIBIRNNANSDDIDLEKKVMUTECEINVASNIDRQ	835
DB	778	EYIFFDSIDNKLKAKSKNIPLGASISEDIKTLLDASVSPDTKFIPLNLKLNIESIGDY	837
QY	836	IVGGRTEAKNLTSDSINYIKNPFKLIBESDSLSYDLKHONGLDDSHFISFEDISKTENG	895
DB	838	IYYEKLEPVKNIITHNSIDDLIDFNLNENYSDELYELKLNLDKYLISFEDISKKNST	897
QY	896	FRIRFINKETGNSIFETEKEIPESEYATHISKEISNIKDTIPDNVANGKLVKVKNLDAAEH	955
DB	898	YSVRFINKNGESVYVETEKEIPEKSYGEHTIKESITIKNSIIDVNGNLNDNLQLDHTSQ	957
QY	956	VNTLNSAFFIQSLIEYNTTKESJNSIVAMKVQVYAQLFSTGANTITDASKVVELVSTAL	1015
DB	958	VNTLNAFFIQSLIDVYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIVDSIQLVNLSNAV	1017
QY	1016	DEIIDLPTLSBGLPIIATIIDGVSIGAAIKELSEYNDPLLRQIEIAKIGIMAVNLTAAS	1075
DB	1018	NDTINVLPITTEGPIVSTLDGINLGAALKELLDEHDLPLKKELEAKVGVLAINWSLSI	1077
QY	1076	TAIVTSALGIASGFSIILVPLAGISAGIPSLVNNELILOQKATKVIDPYKHIASLATEGA	1135
DB	1078	AATVASIVGIGAEVTEIFLLPIAGISAGIPSLVNNELILHDKANSVVNYFNHLSSEKKYGP	1137
QY	1136	FTLDDDXIIMPQDDLVISEIDFNNNNSITLKGCEIWRABEGSGHTELDIDDHFFSPSITY	1195
DB	1138	LKTEDDKILVPIDDLVISEIDFNNNNSIKLGTGNILAMEGGSGHTVTGNIDHFFSPSIS	1197
QY	1196	RKWLSTYDVNLINKEKIDESKOLWLPNAPNRFVGYEMGWTQFPSRLDNDGFKLLDRIR	1255
DB	1198	HIPSLSIYSAIGETENLDFPSKIMMLPNAPSRYFWMETGAVPGLRSLENDGTRLLDSIR	1257
QY	1256	DHYEGQVRYVFAPIADALITKLPRYEDTNVRINLDGNTRSFIVPVITTEQTRKNLSYS	1315
DB	1258	DLYPEKTYWFAFF-DYATTLKPYVEDTNKIKLKDQTNFTIMPITTNERNKLSYS	1316
QY	1316	FYGGSGSYSLSLSPYNNIDNLVENDTWIDVDNVVKNITIESDEIQKGELIENTLSKL	1375
DB	1317	FDGAGGYSLLSSYP-STNINLSKDDLWTFIDNEVRE-SIENGTIKKKGLIKDVLSKI	1376
QY	1376	NIBDNKILNHTTNFVGIDINESNRFLSFLSILEDINIIIEDLVSKSYKILLSGNCKM	1435
DB	1377	DINKNKLIIGNQITDFGSDIDNKORYFLTCELDKDKLSLIEINLVAKSYSLLSGDNK	1436

Qy		1436	LIENSDDIOOKIDHIGFNGEHQKYPISYID-NETKYNGFI DYSKKEGLFTAEFSNESII	1494
Dz		1437	LISNLSNTIEKINTLGID--SKNIAYNTDSENNKIFYGAI-----SKTSOKSII	1483
Qy		1495	RNIYMPPDSNNL-----PIYSSKDIL-KDIRINKGDVKLLIGNFYKD---DMKVSLU	1541
Dz		1484	H-YKDKSNILEFYNDSTLEFNKSDIFIAEDINVMFKODINIITGYYYDNNTDRKSIDS	1541
Qy		1542	FTIEDITNIKLVYLDENGVAQILFPMANAKSALMNTSNSLNFLBSINIKNIFYNNLOP	1601
Dz		1542	ISLVSKNQXVKNGLVYNESVSYLD FVNKSDGHHTNSFMNMLFDLNI SFWKLFGE---	1598
Qy		1602	NIEFILDTNFIISSGNSIQOFELICDKDKNIQYEFNEFKIKTSTYLXYGNRNLIVEPS	1661
Dz		1599	NILNPVIDKITVLGVKNLGVEFICNNKNIDIYFGCEWKTSSKSIIFSGNGENVVVEFI	1658
Qy		1662	YHLDDSGNISSTVINFSQKLYXGIDRYANKVIAPNLXT	1700
Dz		1659	YN-PDTGEDISTGLDFSYPEPLYGIDRYANKVIAPDLXT	1696

RESULT 10

US-08-957-310-6
; Sequence 6, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Kirk, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMEN
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPDH-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-957-310-6

Query Match	51.5%;	Score	4467;	DB	4;	Length	2710;
Best Local Similarity	50.7%;	Pred. No.	2.2e-262;				
Matches	871;	Conservative	366;	Mismatches	440;	Indels	42;
Gaps	15;						
QY	1	MNLVNAQIQKQVYKFRIOEDYVAILNALBEYHNHMESSVVEKYLKLOINNLTDNYL	60				
DB	1	MSLISKEELIKLAY-SIRPRENEYKILTDLDEYNKLTNNENKYVLOQKLNESIDVFM	59				
QY	61	NTYKSGRNKALKKFEYITMVELEKNNSLPVEKNLHFIWIGQINDTAINYNQWKD	120				
DB	60	NYKTSRRNALSNNKKDILKEVILTKNSNTGPVEKNLHFVWIGVEVSDIALEYIKQWAD	119				
QY	121	VNSDYTVKVFYDSNAFLINTLTKTIYESATNTNTLESFRENLDPPDYKNFYKVKMEIYY	180				
DB	120	INAEYNKILWYDEAFVNLTKAIVESSTITLQLEBEIQNPQEDNMKFYKVKMEFIY	179				
QY	181	DQKHFDIYKYSQIEBNEPFIIDNIIKTYLSNEYSKDLBALNKYIEESLNKITTANNGNDI	240				
DB	180	DRQKRFINYKYSQINKPTVPTIDDIKSHLVSEYNRDETVLESYRTNSLRKINSNHGIDI	239				
QY	241	RNLKFADEDLVLYNCELVERPNWLAASDIIRISMLKEDGGVYLDVDILPGIQDPLFKS	300				
DB	240	RANSLFTEOELLNIYSQELNLRGNLAAASDVKLLTALKNFGVYLDVDMLPGHSDLPFT	299				
QY	301	INKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNPDMLEEVORSFESALSKSDKSEIF	360				
DB	300	ISRPSSIGLDRWEMI KLEAIMKYKINIYTSNFPKLDQQLKDNFKLIIESKSEKSEIF	359				
QY	361	LPJDDIKVSPLEVKIAPANNVINQALISLKDYSCLVINQIKNYKILNDNLNPSINE	420				
DB	360	SKLENVNSDLEIKIAPALGVSINQALISKQGSYLTNLVIEQVKNRYQFLNQHLPABIS	419				
QY	421	GTDPTNTMKIPSKLASISNEDNMWMIKITNYLVKGPAPDVRSTINLSGPGVYTCAYOD	480				
DB	420	DNNFTDTTKIFHDSLNSATAENSMPITKIAPVLOQGFMPPEARSTISLSGPGAYASYD	479				
QY	481	LLAFKONSTNIHLEPLRNFEPKTKISOLTEQBITSLWSFNQARAKQFPEYKKGYPE	540				
DB	480	FIMQENITIEKTLKASDLEIEFPENNLSQLTEQELNLSWFPQASAKQFQFYVRYDTG	539				
QY	541	GALGEDNLDFAQNTVLDKDY-VSKTILSS--MKTRNKEYIHYIVOLQGDKISYEASCNL	597				
DB	540	GSLSSEONGVDFNKNTALDKNYLLNNKIPSNVVEAGSNVYHIILOQGGDISYEATCNL	599				
QY	598	FSKDPYSSIIYQKNIGESETAYYYVAD--AEKEIDKTRIPYQISKNKNIKLTFIIGHK	655				
DB	600	FSKNPKNSIIIQRMN--NESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIIGHK	657				
QY	656	SEFNTDTFANLDVDSLSEIETILNLAKADISPKYTEINLLGCNMFPSYSTISAEETPGKL	715				
DB	658	DEFNTSEFARLSVDLSLSNEISSFLDTTKLDISPKNVENLLGCNMFPSYDFNVEETPGKL	717				
QY	716	LLKIKDRVSELMPISODSITVSANQYEVRAINBEGKEIILDHSGKWINKEESIIDKISSK	775				
DB	718	LLSMDALITSLPDVNKNSITIGANQYEVRAINSEGRKELLASHGKWINKEEATMSDLSK	777				
QY	776	EYISFNPKENKIIVKSKYLHELSTLLQEIIRNNANSDDILEKKVMLTECEINVASNIDRQ	835				
DB	778	EYIPFDGIDNKLKAKSNIPCLASISDDIKTLILDASVSPDTKFLNNLKLNISSSIGDY	837				
QY	836	IVEGRIEAKNLATSDSINYIKNEPKLETSDSLSYDLKHONGDSDSHFISFEDISKTENG	895				
DB	838	IYKELFPVKNMIHNSIDDLIDENFLNENVSDELYELKLNLDLDEKYLISFEDISKNNST	897				
QY	896	FRIRFINKEGNSFIETEKEIFSEYATHISKEISNIKDTIFONVNGKLVKKNVLDAAHE	955				
DB	898	YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIIDVNGNLDNLQLDHTSQ	957				
QY	956	VNTLNSAFFIQSLIEYNTTESLNSVMKVQVYAQLFSTGLNTIIDASKVELVSTAL	1015				
DB	958	VNTLNAFFIQSLIDYSSNKDVLNDLSVSVKQVYAQLFSTGLNTIYDVSOLVNLISNAV	1017				

RESULT 11
US-10-011-366-6
Sequence 6, Application US/10011366
Patent No. 6573003
GENERAL INFORMATION:
APPLICANT: Williams, James A.
APPLICANT: Kirk, John A.
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
OF TOXIN A AND TOXIN B FOR THE TREATMENT
DISEASE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/011.366
 FILING DATE: 16-Nov-2001
 CLASSIFICATION: <Unknown>
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US/08/957.310
 FILING DATE: 23-OCT-1997
 APPLICATION NUMBER: US/08/329.154
 FILING DATE: 24-OCT-1994
 APPLICATION NUMBER: US/08/161.907
 FILING DATE: 02-DEC-1993
 APPLICATION NUMBER: US/07/985.321
 FILING DATE: 04-DEC-1992
 APPLICATION NUMBER: US/07/429.791
 FILING DATE: 31-OCT-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Ingolia, Diane E.
 REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPHD-01121
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Query Match 51.5%; Score 4467; DB 4; Length 2710;
 Best Local Similarity 50.7%; Pred. No. 2.2e-262;
 Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;

1 MNLVNAQLOKQVYKFLQEDVEYAILNALAEYHNMSESSVWEKYLKLDNNLTNYL 60
 1 MSLISKEELIKLAY-SIRPRENEYKITLNDVYKNTNNENKYLQKLNESIDVFM 59

61 NTYKSGNKALKKPKYLYLTVMEVLEKKNLSLTPVEKNLHFHWIGQINDTAINTINQWKD 120
 60 NKYKTSNRNALSNLKKILKEVILKNSNTPVEKNLHFHWIGVESDIALEYIKQWAD 119

121 VNSDYTVKVFYDSNAFLINTLKKTTIVESATNNLTLESFPRENLDPEFYKPKRMELIY 180
 120 INAEYNIKLYWSEAFVNTLKAIVESSTTEALQLLEEIQNPQDNKPKYKRMFIY 179

181 DKQKHFIDYKQOISENPEFIDNIKYLSNEYSKDLBALNKYIESLNKITANGNDI 240
 180 DRQKRFINYKQINKPTVPTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSNHGIDI 239

241 RNLEKFADEDLVRLYNOELVERWNLAAASDIILRISMLKEDGGVLDVLDVILPGIQDLPKS 300
 240 RANSLFTEQELNLYSQEQLNGLNLAASDIIVRLALKNFGGVLDVLDVLMPLGIIHSDLPKT 299

301 INKPSITNTSMEMIKLEAIMKYKEYIPGYTSKNFMDLDBEYQRFESALSKSDKSEIF 360
 300 ISRPSIGLDREMIMKLEAIMKYKYINNYTSENFDKLDQDLKDNFKLIESKSEKSEIF 359

361 LPDDIKVSPLEVYKAFANNSVINQALISLKDQSGDLVINQIKRKYILNDNLNPSINE 420
 360 SKLENLVSDLEIKIAFALGVSINQALISKQGSYLTLNVLIEQVKRKYQFLNQHLPAPES 419

421 GTDFNTWKIFSDKLASINEDNMWPMIKITNYLKVGFAPDVRSNTINSGGVYTGAYQD 480
 420 DNNFTDTTKIFHDSLFNSATANSPLTKIAPYLVQGFEMPEARSTISLSGAYASAYD 479

481 LLMFKDNSTNHLLEPELNFPPKTKISQLEQETISLWSFNQARAKSQPEEYKKGFFE 540
 480 FINLQENTTEKTLKASDLIEFKFEPENNLSQLTEQELNSLWSFDQASAKYQEKYVRYDTG 539

541 GALGEDNDLFAQNTVLDKDY-VSKKILSS--MKTRNKEYIHYIVLOQDKITSYPASCNL 597
 540 GSIISEDNGVDNKNLTALDKNLLNNKIPSNVVEEAGSKNYVHYIIQLQDDDISYEATCNL 599

598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKEIDKYRIPYQISNKRNIKLTFIGHGK 655
 600 FSKDPKNSIIIQRM--NESAKSYFLSDGSEIIELNKYRIPERLKNKKEKVKVTFIGHGK 657
 656 SEFNTDTPANLDVDSLSSEIETILNLAKADISPKYIEINLGCNMFSYSISAEITYPKL 715
 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLDISPNKVEVNLGCNMFSDVFNVEETYPKL 717
 716 LKIKDRVSELMPSISQDSITVSANQYVRINEGKREILDHSGKWNKESIIKDISSK 775
 718 LLSIMDKITSTLPDVNKNISITIGANQYVRINSEGRKELLASHGKWNKESIMDSLSK 777
 776 EYISFNPKENKIIVKSKYLHELSTLQEIARNANSSDIDLEKVMLTCEINVASNIDRQ 835
 778 EYIFFSDINLKAQKNIPGLASISEDIKILLDASVPTKFLANLKNLNISSIGDY 837
 836 IVEGRIEERAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDSHFISFEDISKTEG 895
 838 IYVEKLEPVKNIHNSIDDLIDEFNLLENVSDLEYELKLANLDEKYLISFEDISKNST 897
 896 FRIRINKETGNSIFIEFTEKEIFSEYATHISKEISNIKDTTFDNGVNGKLVKKNLDAAE 955
 898 YSVRFINKSGESVYVETEKEIFSEYTHISKEISNIKDTTFDNGVNGKLVKKNLDAAE 957
 956 VNTINSAPFIQSLIEYNTTKESLSNLSVMKVQVYQALFSTGLNTITDASKVZELVSTAL 1015
 958 VNTLNAAFIQLSDIYSSNKNVLDLSTSVKQLYQALFSTGLNTIYDSIQVNLISNAV 1017
 1016 DETIDLPTLSEGLPIATIIDVSGIAAIEKSELSTNDPLLRQETEAIGIMAVNLTAAS 1075
 1018 NDTINVLPTITGIFIVSTILDGINLGAIEKELLEDHPLKKELEAKVGVLAINNSLSI 1077
 1076 TAVTSALGIASGFSILLVPLAGISAGIPSLVNNELIQQDKATKVIDYFKHISLAETGA 1135
 1078 AATVASIVGIGAEVIFLPLPIAGISAGIPSLVNNELIHDKATSVVNVFNHLSKSYGP 1137
 1136 FTLLDKIIMPODDLVLSEIDFNNSNITLGCIEIWRAGSGSHTTDDIDHFFSPSIIY 1195
 1138 LKTEDDKILVPIDDLVISIDFNNSNITLGCIEIWRAGSGSHTTDDIDHFFSPSIIY 1197
 1196 RPEWLSIYDVLAITKKEKIDFSKDLVLPNAPNRVGYEMGTPGFRSLDNDGTCKLLDRIR 1255
 1198 HPSLSIYSAIGIETENLDFSKKIMMLNAPGRVFWETGAVPGLRSLDNDGTCKLLDRIR 1257
 1256 DHEGOFYWRFAFIADALITKLKPYEDTVNRIINDGNTRSFIVPVITTEQIRKNLSYS 1315
 1258 DLYPGFYWRFAFF-DYAITTLKPYEDTVNRIINDGNTRSFIVPVITTEQIRKNLSYS 1316
 1316 FYGSGSYSLSPYNNKIDNLVENDTWIDVNVNKNITTESDEIQGELIENILSKL 1375
 1317 FDGAGGTYSLLSSYPISNTINLSKDDLWIFNIDNEVBREISIENTGKKGKLIKDVLSKI 1376
 1376 NIDNKLILNNHTINFGDINESNRISLTFSILEDINIIIEIDLVSYSKYILLSGNCMK 1435
 1377 DINKNKLIIGNQIDFGSDIDNDRYIFLTCLDDKISLIIEINLVAKSYLLSGDKNY 1436
 1436 LIENSDDIOQIDHIFNGEHOQYIYPSYID-NETKNGFIDYSKKEGLTAEAFSNEI 1494
 1437 LISNLNTIEKINTGLD---SKNIAYNTDSNNKYFGAI-----SKTSQKSI 1483
 1495 RNTYMPDSNNL-----FIYSKDL--KDIRINKGVDKLLIGNYFKD---DMKVSJS 1541
 1484 H--YKDSKNILFEYNDSTLEFNSKDFIAEDINVMKDDINTITGKYVYDNNNTDKSIDFS 1541
 1542 FTIEDNTIKLNGVYLDENGVAQILKFMNNAKALNTSNLWNLFSINIKNIFFYNLDP 1601
 1542 ISLVSKNQVNGVLYLINESVYSYLDLVKNSDGHNTSNFMNLFNLDNISFWKLFGE--- 1598
 1602 NIBFILDNTFIISGNSIQOFELICDKXNIQYFENKIKETSYTLVYGNRNLIVEPS 1661
 1599 NINPVIDKYFTLVGKTNLGYVEFICDNKNKIDIFYGEWKTSSKSTIFSGNRRNVVEPI 1658

QY 1662 YHDDSGNISTVINFESQKLYGDRVKNVIAAPNLYT 1700
 Db 1659 YN-PDGEDISTSLDYSYPLGYIDRYINKVLIAPDLYT 1696

RESULT 12

US-09-084-517-6
 ; Sequence 6, Application US/09084517
 ; Patent No. 6613329
 ; GENERAL INFORMATION:
 ; APPLICANT: KINK, JOHN A.
 ; APPLICANT: WILLIAMS, JAMES A.
 ; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
 ; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200
 ; CITY: SAN FRANCISCO
 ; STATE: CALIFORNIA
 ; COUNTRY: UNITED STATES OF AMERICA
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/084,517
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/
 ; FILING DATE: 16-MAR-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/329,154
 ; FILING DATE: 25-OCT-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/161,907
 ; FILING DATE: 02-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/985,321
 ; FILING DATE: 04-DEC-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/429,791
 ; FILING DATE: 31-OCT-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: CARROLL, PETER G.
 ; REGISTRATION NUMBER: 32,837
 ; REFERENCE/DOCKET NUMBER: OPD-01610
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 705-8410
 ; TELEFAX: (415) 997-8338
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2710 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-084-517-6

Query Match 51.5%; Score 4467; DB 4; Length 2710;
 Best Local Similarity 50.7%; Pred. No. 2.2e-262;
 Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;
 QY 1 MNLVKAQLQKQVYKFRIGEDYVAINALBEEYHNSSSVYKYLKDKDINNLDNYL 60
 Db 1 MSLSKKEELIKLAY-SIRPRENEYKTILTNLDEYKNKLTNNENKYLQLKKNESIDVFM 59
 QY 61 NTKYKSGRNKALKFKPKYLTVEVLELKNKSLTPVEKONLHFINTGGQINDTAINYNQWKD 120
 Db 60 NKYKTSRRNALSNNKDKILKEVILIKNSNTSPVEKONLHFINTGGQINDTAINYNQWKD 119

QY 121 VNSDYTVKVFYDSNAFLINTLTKTIIVESATNTTLESFRENLDPEFDYKFKRMEIY 180
 Db 120 INAEYNIKLAWDSEAFVNTLKAIVESSITTEALQLLEEIEIQNPQDMFKFYKRMFIY 179
 QY 181 DKQKFTDYKYSQIEENPEFIIDNIITYLSNEYSKOLEALNKYIEBSLANKITANNNDI 240
 Db 180 DRQKRFINYKYSQINKPTVPTIDDIIMSHLVSEYNDRDETIVLESYRTNLSLRKINSHGIDI 239
 QY 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGLOPDLFKS 300
 Db 240 RANSLFTEQELLNTYSQELNLRGNAASDILVLLALNFGVYLDVDMPLGHSOLFKT 299
 QY 301 INKPDSTINTSWEMIKLEAIMKYEIPGYTSKNFMDLDEBEVORSFESALSSEKSEIF 360
 Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYNNTYSENFKLDQDLKDNFKLIIESKSEKSEIF 359
 QY 361 LPLDDIKVSPLEVKIAPANNVINQALISKDSYCSOLVINQIKRKYILNNDLNPSINE 420
 Db 360 SKLENLVSDLEIKIAFALGVSINQALISKGSYLTNLVIEQVKNRYQFLNQLNPAIES 419
 QY 421 GTDFTNTMKIPFSDKLASINEDNNMFMKITYLKVGFADPVRSTINLSGGVYTYGQD 480
 Db 420 DNNFTDTTKIFHDSLFNSATAENSMTLKLAPYLVQVGFMPPEARSTISLSGPGAYASYD 479
 QY 481 LLMEKDNSTNIHLEPELRNFERPKTKISOLTEQETISLWSFNQARAKSOFEEYKKGFFE 540
 Db 480 FINQENTIEKTLKASDLIEFKFPENNLSQTEQINSLSWFOQASAKYQFEKTVRYTG 539
 QY 541 CALGEDDNLDPQNTVLKDY-VSKILSS--MKTRNKEYIHYIVQLQGDKISYEASCNL 597
 Db 540 GSLSEDNGVDNFKNTALDKVLLNKNIPSNVVEAGSKNYVHYIIQLQGGDISYEATCNL 599
 QY 598 FSKDPYSILYQKNIEGSETAYYYVAD--AEIKEDKIRIPYQISKNRIKLTFFIGHGK 655
 Db 600 PSKNPKNSIIIRNM--NESAKSYFLSDGSESILELNKYRIPERLKNKKEKVKVFFIGHGK 657
 QY 656 SEFNTDTFANLDVDSLSEIETILNLAADISPKYIINLLGCMFYSYSIAEITYPKL 715
 Db 658 DEFNTSEFARLSVDSLSEISFSDTKLDISPKNVEVNLGCMFYSYDFNVEITYPKL 717
 QY 716 LLKIKDRVSELMPISIDSTVSANQVEVRINEGKEIILDSGKWKINKEESIYKDISSK 775
 Db 718 LLSIMDKITSLPDVKNKSITIGANQVEVRINSEGRKELLAHSGKWKINKEEAINSDLSK 777
 QY 776 EYISFPNPKENKIIIVKSKYLHELSTLQEIIRNANSSDILEKKVMTLCEINVASNIDRQ 835
 Db 778 EYIFPDSIDNKLKAKSKNIPGLASISEDITKLLLDASVSDPTKFIILNKLNIESSIGDY 837
 QY 836 IVEGRIBEAKNLTSDSINYKNBFKLIESDSYDLKHQGLDSDSHIFSPEDISKTENG 895
 Db 838 IYVEKLEPVKNIIEHNSIDDLIDEPNLENVSDLEYELKLNLDKYLISPEDISKNST 897
 QY 896 FRIRFINKETGNSIFITEKEIFSEYATHISKEISNIKOTIFDNVNGKLVKKNLDAHE 955
 Db 898 YSVRFINKSNGESVYVETEKEIFSKYSEHTKEISTKNSIITDVNGLLDNIOQDHTSQ 957
 QY 956 VNTLNSAFFIOSLIETYNNTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTAL 1015
 Db 958 VNTLNAFFIOSLIDYSSNKDVLDNLSTSVKQVYAQLFSTGLNTIYDSICLVNLISAV 1017
 QY 1016 DETIDLLPTLSEGLPIIATIDGVSGLGAAIKELSETNDPLLRQRIEAKIGIMAVNLTAAS 1075
 Db 1018 NDTINVLPTITEGPIVSTILDGINLGAAIKELDEHDPLLKKELEAKGVGLAINMLSI 1077
 QY 1076 TAITVSAIGIASGSPSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISIAETGA 1135
 Db 1078 AATVASIVGIAEVTIFLLPIAGISAGIPSVNNELIHDKATSVVNYFNHLSSESKYGP 1137
 QY 1136 FTLDLDDKTIIPQDDLVISEIDFNNNSTLKGCEIWRAGSGSHTLTDDIDHFFSSPSITY 1195
 Db 1138 LKTEDDKILVPIDDLVISEIDFNNNISKIGTCNILAMEGGSGHVTGIDHFFSSPSISS 1197
 QY 1196 RKPWLSIYDVNLNIKKEIDFSDKOLMVLNPNAPRVFVEMGWTPGFRSLDNDGTLLDRIR 1255

Db 943 VLRSKVD---SYINEN-----KQNLKEDA--GALRDYAKLKKMKELFSLMLDDGKYK 990
Qy 736 TVSANQYVRINEGKREILDHSGKWINKKEEIIKDISKEVISFNP-----K 783
Db 991 IITNAY---IKERDL-----SGIINYIENSIHGESFDIIRSNQHEWGLSTVEQFK 1042
Qy 784 ENKIIVKSKYLHSLTLOEIRNANSDDIDLEKKYMLTECEINVAS-----NIDRQIVE 838
Db 1043 KFEFYKSE--LSAKSIFDDIKNK-YITDPETKRNVLVHQLSDIKERIAFLDISHYAVP 1100
Qy 839 GRTEAKNLTSDSINIKNEFKLIESI-----SDSLYDLKHQNL 878
Db 1101 GSLEKLQS-----GVFSDINIIAAYLLASVGSYHGVVPAPSKLLELRHTK 1156
Qy 879 DSHFISFEDISKTEGFRIRINKETGMSI---FIETEKEIFSEYATHISKEISNIKT 935
Db 1157 SNEWI--EKITP---YVYDILSDNVNVRPLSEEQKILNDIKLEISKVS---EQ 1207
Qy 936 IFNVNGKLVKNLDAHEVNTLNSAFIQLSLEYNTTK-----ESLSNLSV----- 983
Db 1208 YFMKLTQKSSVIGIKYSVDFDYRNLFLSLPINQNLTPPMRYFEMLYDIHIGIEN 1267
Qy 984 -AMKVQVYACLFTGLNTIDASKVVELV-----STAL-DETIDL 1021
Db 1268 KAREFYKSFSSUNLDFLINDERVNLNLEGLIKKYKLSLSEIHRTLTSTSPADISPL 1327
Qy 1022 LPTLSELPLIATII-----DGVSLGAIAK-----ELSETNDPLLR 1057
Db 1328 LQNI---CPSITIIKKTEYVGHQLTNMTVASVVKPYDFSNLGAINDSKSVDPALH 1384
Qy 1058 QEIB-AKIGIMAVNLTAASAIVTSAL-----GIASGPSILL 1093
Db 1385 TIVEQAKYNLLSNWDFYNTHASITWDTARQHKSTNIEFHPQSLDFDRDQSGKGLGLSLY 1444
Qy 1094 VPLAGISAGIPLSNV---ELILODKATKVIDYFKHISLAETEGATLLDXKILMPQDD 1149
Db 1445 LDTGGYGGYQKLRHNTDASTLYQTK-----YNDNLKLSNRDDPFLKRTORIITMSNE 1498
Qy 1150 LVLSEIDFNNSITLGCETWRAEGSGHGLTDDI--DHFFSPSITYRKPLMSIYDVNL 1207
Db 1499 L-----GNRLNKAOLEVLEK---DPILTGLIYORRISLITTEYHSLALQOISS 1548
Qy 1208 IKK-----EKIDFSKDLNVLNPNRNVFGYEMGWTFGFSLNDGDKJLD 1252
Db 1549 FWRVTDNFGHCDPHSLAQAALTIKNI-----TSNR-----NFSSLYSGSI--- 1589
Qy 1253 RIRDPH-EGOFYWRFYAFIADALITKUPRYEDTNVRINDG---NTRSPFVPIVTEQI 1308
Db 1590 -VKIYFSESLLNNWKYIKLPIVQTGSLRLDIYLTPEKLSSTGGSLLIMGLHLEV----- 1642
Qy 1309 RKNLSYGF-YGSGGSYLSLSPYNNMIDNLVENDTWIDVNDVKNITIESDEIQ--- 1363
Db 1643 -----SFIYDIGV-----INGNRIBEST---DVGNKIRSLKINGDILQHYIN 1682
Qy 1364 -----KGELIENILSKLNTEDNKI-----ILN---NHTINFY- 1392
Db 1583 THYLSSEQTQKIDIVDFLQDNTIKVKLESIDKPISEIQOPLHLSILSKQEHVKNLLS 1742
Qy 1393 GDINE-SNRF-----ISLTFSEILEDINI-----IIEI---DLVSKSYKILLSGNCKML 1436
Db 1743 GLLDFFSNKLRKQGLSLKTNVLSVNNFKESKINSDDTVETVTDLQGRLYRVDIDTRVIGL 1802
Qy 1437 I--ENSSDIQKIDHIFNGEHQKPIYPIYIDNETKNGFIDYSK-----KEGLFTABFSN 1490
Db 1803 TFKEGINSLSALEHNM-----IDAIMSVIGLVQVYARMFKMNDNISADHAG 1849
Qy 1491 E-SIIRNIYMPDSNNLFIYSSDKIDIRIINKGVK-LLIGNYFKDDMKVLSLFTIEDTN 1548
Db 1850 AVSDIKNIY-----DKFLGGILITTNRNVNPGVSGASLEGFISSGLEVCAAS----- 1897
Qy 1549 TIKLNGV---YLDENGVAQILKFMNNAKSALNTSNSLNFLESTINIKNIFFNNLD----- 1600
Db 1898 --RMGGTAGRYL--SNVAKVYKL-----PLLDI--GLINWSLYDSSLSINAKAT 1939

Qy 1601 PNIEFI---LDTNF-IISGNSIGQPE---LICDCKNIQPIYFINKIKETSVTLYV-- 1650
Db 1940 TQBEYIITADIVSPSSINTALSICATAYPELAIA-----IVPITI-FSHEVKYAVYVNO 1993
Qy 1651 -GNRQNLIVFSPVHLDD-----SGNISSTVINFSQKLYG---IDRYVNVKVI 1694
Db 1994 INERHKLWLEAKVLDNGSAKVLISINKATGIIDLSNNQVLGNLYLDRENPPL 2047

RESULT 14

US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 4.0%; Score 347; DB 4; Length 10182;
Best Local Similarity 19.4%; Pred. No. 2.5e-11;
Matches 354; Conservative 318; Mismatches 662; Indels 494; Gaps 97;

Qy 20 QDEYVAILNALEYHNMSESVVEKYLKLDINN-----LTDN-----YLN-- 61
Db 7557 QKD---TILNHI---FSAPTRSQVGEKIASAKQLNNTMKALRDSIADNNEILQSSKYFNE 7611
Qy 62 TYKSGRKNALKKPEYLTWELKKNLSLTPV-----NDQPTFVMADEIQSVLNEVKQTKDLN 105
Db 7612 SEQQAYNAVQNAKKNII-----NDQPTFVMADEIQSVLNEVKQTKDLN 7659
Qy 106 Q--IND-----TAINVINQWKDVNSDYTVKYFYDSN-----AFLINTLKKT 144
Db 7660 QKLANDKTDQAATNALNLYNQAGNLETKVQ---NSNSEPEYQKVQVQLANQLNDAMKK 7716
Qy 145 IVESATNNLTSPRENLDPEFDYKFKYRKMELIYDKQKHFDIYYSQIEE--NPEF-- 200
Db 7717 LDDALITGN--DAIKQTSVINEITSQ-----QVNFDE---YTDGRKNIVAEQTPNNSP 7765
Qy 201 -----IIDNIIKTVLSNEYSKDLEALNKY-----IBESLNKITANNNGDIRNLEKFADEDLV 252
Db 7766 TNINTIADKI-----TEAKNDLHGQVKLQAOQSINTI-----NQMTGLNQAKQEL- 7813
Qy 253 RLYNDELVERNLAASDILRISMLKEDGGVYLDVILPGIQPLPKSINKPDSITNTSW 312
Db 7814 ---NOEIQOTQTRSEVHGVINKAQAALNDS-----MNTLRQSTIDEHEVKQTSN 7858
Qy 313 EMIKL-----EAIMKYKEYIPIGTSKPNFMDLDEEVQSFESALSKSDSKSIFLPLD 364
Db 7859 YINETVGNQYANNNAVDKVKQIINQTSNFTWNPL--EVERATSNVKSJL-K-DALHGERELN 7915
Qy 365 DIKVSPLVEVKAFANNSV--INQALISLKDYSYCDL-----VINQIKNKRYKILN----- 411
Db 7916 DNKNSK-----TFAVNHLDNLQA---QKEALTHEIQATIVSQVNNIYNKAKALNNDMK 7967
Qy 412 -----DNLNPSINEGTDFNTMKLFPKSLASISNEDNMFWI-----KITNYL 454
Db 7968 KLDIVAQODNVQRQSNYINEDSTPQNNYND---TINHAQSIIDQVANTPTWSHDEIENAI 8024
Qy 455 KYGFAPDVRSTIN--LSGPGVVTGAVQDILLMFKDNSTNIHLEPELRNPEFF-KTKISQLT 512

Db 8025 N-----NIKHAINALDGE-----HKLOQAKENA-----NLLINSINDLNAPORAINELV 8069
 QY 513 E-----QBITSLWFSNQA-----RAKQF-----EYKQYFEGALGED 546
 Db 8070 NEAOTREKVAEQLQSAQALNDAMKHLRNSIQNQSVRQESKYINASDAKQYHNAVRE 8129
 QY 547 DNLDAQNTVLDKDYVSKILSSMTKRYEYHYIYVLOQDKIYEASCN--LFSKDPYS 604
 Db 8130 ENINEQHTLDEKII-KQLTDGVNQAND-LNGVELLDADQNAHOSIPTLHNLQAQQ 8187
 QY 605 SILYOK-NIEGSETAYYYVADAIEKIDKVPYQISKNENIKLTFIGHGKSEFNTDF 663
 Db 8188 NALNEKINNNAVTRTEVAALIGQALKDLDHAMENLEBSIKDQVK-----QSNVINEDSD 8242
 QY 664 ANLDVDSLSSETETILN-----LAKADISPKYIEIN-----LLGCMNFSYSISAEETYP 712
 Db 8243 VQETVDNAVDHVTETILNVTNPTLSIEDIEHAINEVNOAKQLRGKQKLYQTDILAD--- 8299
 QY 713 GKLLIKIDRVSELMPSISQDSITVSANQYEVINEBGR-----EILD 756
 Db 8300 -KELSKLDLTQSSSSISNQIYTAKTREVAQAIEKAKSLNHAKMLNKYVQADKYL 8358
 QY 757 HSGKWINK-----ESTIKDISKEYISFNPKENKIIKSKYLHELSTLLQIRNN-- 807
 Db 8359 -SSRFINEQPPEKAYQQAINEVDVSIHRQTNPEMDPTVINS-ITHELETAQNHLHGQK 8416
 QY 808 -----ANSDIDLEKKVMLTECEIN-----VASNIDR-QIVEGRISBEAKNLT 848
 Db 8417 LAHAQDAANVINGLIHLNVAQREVMINTNTATTREKVAKULDNAQALDKAMETLOQVV 8476
 QY 849 SDSINYKNEFKLISDISLDYLDKHQGLDSDSHFISFEDISKTEGFRIFINKETGNS 908
 Db 8477 AHK-NNILNDSKYLN--EDSKYQOQYDRVIADAEQLLNQNTNPTLEPKYKVIDKQNV--- 8530
 QY 909 IETETEKEIFSYVATHISKEISNIKDTIFDQVNGKLVKKVNLDAHAHEVNTLNSAFFQSL 968
 Db 8531 --LANEKILFG--AEKLSYDKSNANDEI-----KHVNYLNA----- 8563
 QY 969 IEVNTTKESLNSLVAMKQV---VVAQLPSTGLNTITDASKVVLVSTAL----- 1015
 Db 8564 -QKQSIKDMISHAALRTEVKQLLOQAKILDEAMKSLDEKTVQV-ITDITLPHYTEASEDK 8621
 QY 1016 DETIDLPLTSLGFLFIATII-DGVSLG---AAIKELSETNDPL---LRQIEAKI----- 1064
 Db 8622 KEKVD--QIVSHAQALIDKINGSNVSLOQVQALBQLTQASENLDDQORVB-EAKVHANQ 8678
 QY 1065 ---GIMAVNLTRAFTA--IVTSALGIASGFSILLVPLAGISAGISLVNELLIOQKATK 1119
 Db 8679 TIDQLTHLSLOQOTAKESVKNATKLEE-----IATVSNNQAALNKVMGK 8723
 QY 1120 VIDYFKHISLAETGAFTLLDDKIIMPQDVLVSEIDFNNSITLKGCEIWEAGSGHT 1179
 Db 8724 LEQFINHADSVDNSNYROADDKIAYDEALEHGQDIQKTNATQNETK----- 8772
 QY 1180 LTDDIDHFFSPSITYRKELWLSIYDVLN-IKKEKIDFSDKMLVLPNAPNVPFGVMGWTP 1238
 Db 8773 -----QALQOLIVAFETSLNGFELNHPARPALEYIKSLKINNQAOKSALEDKVTQSH 8824
 QY 1239 GFRSLD---NDGFKLDRIRDHYEQFYVRYPAFTADALITKLKPRYEDTNVRINLDGNT 1295
 Db 8825 DLELEHIVNEGTLNDIMGE-----LANAIVN---NYAPTAKASIN----- 8862
 QY 1296 RSFIVPVITTEQIRK-NLSYSYFGGGYSLSLSPYNNMIDNLVENDTWIDVD----- 1349
 Db 8863 -----YINADRLKDNFTQANNAFDALN---KTQGNLDFNAI--DTFKDDIFPKTXDA 8911
 QY 1350 -NVVKNITTESBIOKGLIEMI--LSKLNI--EDNKIILNHTINFGYDINESNRFISLT 1405
 Db 8912 LMGIELTAAKKAER--LIDSLEKINKAQFTHANDEINNTNS-----IQLSRIVNQA 8963
 QY 1406 FSILEDINIITIEDLVSKYKILLSGNCMKLIENSDDIOQKIDHIGFNGEHOKYIPYSYI 1465

Db 8964 FD-LNDAMKSLRDELNNQAFVQASSN---YINSEDLKQFDFHALSNA--RKVL----- 9012
 QY 1466 DNETKNG-PIDYSKEGLFTAEFNSIESIIRNIYMPDSNNLFIYSSKD-LKDIRINKGD 1523
 Db 9013 ---AKENGKLDKQIOGL-----KQVIEDTKDALNGIQRLSKAK 9049
 QY 1524 VKLLIGNYFKDDMKVLSFTIEDTNTIKLNGVYLDENGVAQILKFMNNAKSAINTSNLSIM 1583
 Db 9050 AKAI--QYVQ-----SLSYINDAQRHIAENNH-NSDDLSSLANTLSKASDLDNAKDLR 9101
 QY 1584 NFLESINI---KNIFYNNLDNPIEFILD 1608
 Db 9102 DTIESNSTSVPSNVYINADKQLQIEFD 9129
 RESULT 15
 US-09-543-681A-4980
 ; Sequence 4980, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709,1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; PRIOR FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 4980
 ; LENGTH: 2807
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-4980
 Query Match 3.9%; Score 334; DB 4; Length 2807;
 Best Local Similarity 17.5%; Pred. No. 2,4e-11;
 Matches 354; Conservative 311; Mismatches 661; Indels 694; Gaps 86;
 QY 2 NLVNAQLOQVWVVKPRIQEDVEYVAILNALBEEVHNS---ESSVVEKYLKLDINNLTN 58
 Db 861 NMTADELSAAEVTGKARGENYKAIQSLRDYELQSTNTNTPIDIFQGLSTLRQINS 920
 QY 59 YLNTYKSGRNKALKKFKYLWME-----VLEKNNSLTPVEK 96
 Db 921 YLLKHPDSKNEALQQLRQVDIRYKHASILLNSQVDSNNNTFSYLYEIFSANLKNK 980
 QY 97 NLHFIWIGGINDTAINYNQWQVDSYTVKVFYDSNAPLINTLKTIVE-----SATNN 152
 Db 981 HIYL-----DENGFTVTKGN-NLQNDKLISGNA--ITSIKRLVEYGESEITNQ 1029
 QY 153 TLESERENLDPEDYN-----KFKYKME-----IYDKQKH-----FI 187
 Db 1030 VFNQFTN---EFAQNGHIDLSGLKTHQAEQQVSISSSTLYTWKSKSRGLHAAL 1085
 QY 188 DYTQSGIENPERIITNIKTVLS-----NEYSKDLEA---LNKYIEE 227
 Db 1086 QICSGRLQSDSDVDHNNYVSWPAGSKSLANPLNITTEKNPDLKLRYDLSQVSR 1145
 QY 228 SLN-----KITANNQNDIRNLEKFADEDLVRLYNOELVERWNLAASDI 271
 Db 1146 STNFMT-LQIDIOBEKTNFKLNKDEGNDRLKPEFD-----KTNLTRGLE- 1191
 QY 272 LRISMLKDEGGVYLDVD-----ILPGIQLPDLKXINKPD-----SITNTSWEM 314
 Db 1192 ---NITPETAGLLNAPHWIPETLIPPEISHPFIKQWSPNLSNLEVSNTFSAVKNWLK 1248
 QY 315 IKLEATMKYKEYIPGYTSKNFMDLDEVOQSPESSALSSKSKSEIF-LPLDDIKVSPLEV 373
 Db 1249 VAPPTAMINGTF--ALEIKKIEIRKENTENKTSVADDTNDGVPRINLEGLDAAAMQK 1306
 QY 374 KIAFANNVINOALISLKDSYCSDLVINOIK----- 404

Db 1307 EWKINSQ--SDPRYQLTKNCSSIVARILKAGGADQLIGHHMKPRFGIWTNPENLYKFSQ 1364
QY 405 -----NRYKILNDNLNPSINEGDTDFNTTKIFSDK-L 435
Db 1365 KIQEARLAQIAVOQCKPIHNSLOALSDHNKVAINDNCTSP--NERKNLSPLTRFFNDHFF 1422
QY 436 ASISNEDNMFMKITHYLVKVPAPDV-RST----- 465
Db 1423 GSYERRRDMTVMRKNEEKITLKGDAGRUTGSYRGNDNIPBATTCKVVLFLHGSNSPT 1482
QY 466 -----INLSGPGVYTGAYQDILLMFKDNSTNIHL-----BPE-- 497
Db 1483 EKQSSSFHYNNQOGIDMLAINRFGESDGSPTQOQYADACTMFRYLVDNKGIDPKNI 1542
QY 498 -LRNFEPKTKISOLTE-----QETSLW-----SFNQAPAKSQPEYKKGYPFEGALGED 546
Db 1543 IHHGYSNGAPITAAKLAISANGQHIAGLFIDRPMSPMSKAIESYDEYSLKLTQLAKK 1602
QY 547 DNLDFP-----QNTVLDKDYVSK-----KILSSM 570
Db 1603 INGQFSVEKNLQGFSDKIPILLTDCGELGIGGEKMTKLLDKGYLVKGBERTDVSHLSL 1662
QY 571 KTRNK-----EYTH----- 579
Db 1663 KLMQYKQIISLSLSPYVNTNHRNALDSIQNPICKNKVTDGWMYVVPNDNGDSRF 1722
QY 580 ---YIVOLQGDKISYEASCLFSDPYSSILYQKNIEGSETAYYYYVADAEIKEIDKYRI 636
Db 1723 DSQIIQTEDDPIVAQASALAAKXNSIVQLDADG-----HYRVA-----YGD 1768
QY 637 PYQISNKRNIKLPFIGHK--SEFNTDTFANLDVDSLSSEI---ETILNLA-KADISPKY 690
Db 1769 PAQLSGK--IRWLQVCHGRETTENNHLRSLNNADELSQLVKFNTVFSKNNINTPEH 1826
QY 691 IEINLGCNMFYSISABEYTPGKLLIKIDRVSELMPSISQDSI--TVSANGYEVRIE 748
Db 1827 ISI--VGCSTISNDKQAGFAH-----QFINALDQOQIRSSVSARVTEVAVDS 1871
QY 749 EGKEILLDHSKWKNEESIIOKDISKEYISFNPENKNIIVKSKYLHELSTLLQETRNNA 808
Db 1872 NGHYTKDNNGEWAKO-----NQNKVNVNNKKGKITTEFEQIUNG 1914
QY 809 NSSDIDLEK-----KVMLTECEINVASNIDQIVGRIEEAKNLTSDSINYKN--- 857
Db 1915 AEGDIDLTKIGASADNKRVGAIADNNEVFTAPKKGKSTKQSSSGSDNSLSYSYGNIOVS 1974
QY 858 -----EFKLIESISLSL-----YDLKHQGLDDSHFISFED 888
Db 1975 VGDGEFTVNWGTSNGLIKVGTGFKSLVFGDNNVMVHVGNGDSKH--SVDIAGYQAFEG 2032
QY 889 ISKTENGFRIPFINKETGNSIFITEKEIFS-----EYATHISKEISNIKDTIFDNVN 941
Db 2033 MQLFVGTRNVSF-NLGRSNDLIVMLEKSIPPTPLINFPDGAARITKVLESIAACSGENNDN 2091
QY 942 GKL-----VKNVNLDAAEVNTLNSAPFIQSLIEVN-----TTKESLSNLSVAM 985
Db 2092 DWLSVQNDQWTLGAKKFAIDNSGIDQTSWDY--KTLTDLNSQNKSSRGLKSDLEATL 2149
QY 986 KVQVYQALFSTGLAT---ITDASKVVELVS-TALDETIDLLPTLSEGLPIIATIID-GV 1039
Db 2150 NKKYNQWLGSGNNENGKISRAEQRLQDLSKIAFNFAIG-----GQGADIQVTAGNWL 2204
QY 1040 SLGAIAKELSTNDPLURQETEAKIGIMAVNLTA---ASTAIVTSALGIASGF-SILLVP 1095
Db 2205 MFGDNIQSIIDTN-----LGSFLGLMTQOFTATGMVETFTYTPQNPLTKLNKLLGR 2257
QY 1096 LAGISA-----GIPSLVNNEL-----ILQD-----KAT 1118
Db 2258 LASVSSDRTLGBIFGVDSTTGKTIISRNGEPIDVVAILKOMLAVVTEFGGEKLSAFTDPT 2317
QY 1119 KVIVDYFKH-----ISLAETGAFLLDDKIMPODDLVLSEIDFN-NNSITLGKCE 1168
Db 2318 KLLDNLQSSINLGKGIISFAESHG-----LXK--ATEDQQNESEVSINACTPITTSQTE 2371

QY 1169 IWRAGGSGSHLTDDIDHFRSSPSITYRKPMWSIV-----DVLN 1207
Db 2372 -----NSKITEENRPPFGFNSLNIPNLPATIFNKDKQTDMDRLAENLKENLAADVLN 2423
QY 1208 IKKEKIDFSK-----DLMWLPNAPNRVFGYEMGWTPGFRSLDNDGTKLLDRIRDHY 1258
Db 2424 MEKTLDFLRNSGHLKGDGDHIV-----SLGNYNFNW-----GGDNDLGAYLGDN 2470
QY 1259 EGQFYW-----RYFAFIADALIT-----KUKPRYEDTNVRINLDGNTSRFIVPVITTEQ 1307
Db 2471 N---FWGGRGSDTFFATGISNIFTGGYNDLGVLGRSNMFMFGGDGDDTALIAGRINNVY 2527
QY 1308 IRKNLSYSF-YGSGG-----SYSLSPYNNIDNLNLTVENDTWIDVNVVKNITIE 1358
Db 2528 LGDGLDKAFVFGEGGEIHTNAGNDYAVTTGNYNRFSGS--EQD-FVVTIGN-HNQISLE 2583
QY 1359 -----SDEIQ-----KGELIENILSKL-----NIBDN 1380
Db 2584 BONDPAKIFGNVNRNLNGGGGDEIQLMGYHAVVGGDNDQJIAASFSLNLTNGTND 2643
QY 1381 KIILNNHTINFGDINESNRFIISLTFSILEDINIIE--IDLVSXSKILLSG-NCMKLI 1437
Db 2644 IILGGYQNHFKGGNGIDSLIN-----NNVIDCYIDDISQEDNIVLGDDIDWNKLM 2694
QY 1438 ENSDIOQKIDHIGFNGEHQKVIPIYSYIDNETKINGFIDY 1477
Db 2695 PERSGYDLKISHIRPPQNTGQATFERIGSTT-----FTDY 2730

Search completed: April 1, 2004, 16:46:06
Job time : 35.625 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:33:50 ; Search time 71.25 seconds

(without alignments)
6741.484 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700

Perfect score: 8675

Sequence: 1 MVLVNAQLQKVVYKFRIQ.....VLYGIDRYVKNVLIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6756.5	77.9	2366	2 AAR95011	Aar95011 C. diffi
2	6756.5	77.9	2366	2 AAW68388	Aaw68388 Clostridi
3	4467	51.5	2710	2 AAR95016	Aar95016 C. diffi
4	4467	51.5	2710	2 AAW68387	Aaw68387 Clostridi
5	2792	32.2	546	2 AAY33700	Aay33700 C. sordel
6	512.5	5.9	3169	6 ABO14444	Abol14444 E. coli 0
7	431.5	5.0	4688	6 ABU48941	Abu48941 Protein e
8	398	4.6	1979	3 AAB18171	Abal18171 Plasmodiu
9	374.5	4.3	3043	6 ABM68924	Abm68924 Photorhab
10	372	4.3	2410	6 ABM68965	Abm68965 Photorhab
11	355.5	4.1	5024	4 AAG82935	Aag82935 S. epider
12	347	4.0	6641	6 ABU42656	Abu42656 Protein e
13	347	4.0	10182	5 ABP38314	Abp38314 Staphyloc
14	343	4.0	2485	3 AAB18172	Abal18172 Plasmodiu
15	340.5	3.9	5005	6 ABU48947	Abu48947 Protein e
16	336	3.9	4620	6 ABU40997	Abu40997 Protein e
17	313.5	3.6	3973	3 AAB18253	Abal18253 Plasmodiu
18	312	3.6	1639	2 AAW54145	Aaw54145 P. falcip
19	312	3.6	1639	5 AAE29345	Aae29345 Plasmodiu
20	311	3.6	1516	3 AAB18195	Abal18195 Plasmodiu
21	305.5	3.5	2496	3 AAB18222	Abal18222 Plasmodiu
22	304.5	3.5	1802	3 AAB18217	Abal18217 Plasmodiu
23	304	3.5	2184	4 AAE00425	Aae00425 P. falcip
24	302.5	3.5	1254	2 AAR07503	Aar07503 Merozoite
25	302.5	3.5	1254	2 AAW24575	Aaw24575 Merozoite

ALIGNMENTS

RESULT 1

AAR95011
ID AAR95011 standard; protein; 2366 AA.

XX	AC	AAR95011;			
XX	AC	AAR95011;			
DT	16-OCT-2003	(revised)			
DT	08-JUL-1996	(first entry)			
XX					
DE					
XX					
XX					
KW					
KW					
XX					
OS					
XX					
PN					
XX					
PD					
XX					
PF					
XX					
PR					
PR					
PR					
XX					
PA					
XX					
PI					
PI					
XX					
DR					
DR					
XX					
XX					
PT					
PT					
PT					
XX					
XX					
CC					
CC					
CC					
CC					
CC					
CC					

Abm71130 Staphyloc
Abal18198 Plasmodiu
Abal18167 Plasmodiu
Aap50777 Sequence
Abal18250 Plasmodiu
Abal18180 Plasmodiu
Abu25521 Protein e
Abu44210 Protein e
Aaw22017 Utraphin.
Abr53560 Protein s
Abr52698 Protein s
Acb46133 Plasmodiu
Abu44208 Protein e
Abj19025 Pathogen
Abp73774 Candida a
Abal18218 Plasmodiu
Aap73809 Candida a
Abu24813 Protein e
Abu43311 Protein e
Aab18199 Plasmodiu

26 302.5 3.5 1370 6 ABM71130
27 298.5 3.4 2539 3 AAB18198
28 296.5 3.4 1308 3 AAB18167
29 296 3.4 1654 1 AAP50777
30 293.5 3.4 1844 3 AAB18250
31 291 3.4 2295 3 AAB18180
32 290 3.3 1182 6 ABU25521
33 289 3.3 2575 3 ABU44210
34 289 3.3 3433 2 AAW22017
35 288 3.3 1875 6 ABR53560
36 288 3.3 2000 6 ABR52698
37 287 3.3 1610 7 ADB46133
38 286.5 3.3 2046 6 ABU44208
39 286.5 3.3 2681 6 ABJ19025
40 285.5 3.3 1948 5 ABP73774
41 285 3.3 2010 3 AAB18218
42 283.5 3.3 1881 5 ABP73809
43 281.5 3.2 1196 6 ABU24813
44 281 3.2 3692 6 ABU43311
45 280 3.2 1436 3 AAB18199

Abm71130 Staphyloc
Abal18198 Plasmodiu
Abal18167 Plasmodiu
Aap50777 Sequence
Abal18250 Plasmodiu
Abal18180 Plasmodiu
Abu25521 Protein e
Abu44210 Protein e
Aaw22017 Utraphin.
Abr53560 Protein s
Abr52698 Protein s
Acb46133 Plasmodiu
Abu44208 Protein e
Abj19025 Pathogen
Abp73774 Candida a
Abal18218 Plasmodiu
Aap73809 Candida a
Abu24813 Protein e
Abu43311 Protein e
Aab18199 Plasmodiu

CC protein, are used to raise avian antibodies useful as antitoxins or
CC diagnostics, and in vaccine prodn. (Updated on 16-OCT-2003 to standardise
CC OS field)

SQ Sequence 2366 AA;

Query Match 77.9%; Score 6756.5; DB 2; Length 2366;
Best Local Similarity 76.5%; Pred. No. 0;
Matches 1301; Conservative 192; Mismatches 206; Indels 1; Gaps 1;

Qy	1	MLVYKAOIQKQVYKFRIOQDEYVALINLALEYHNHMSSESSVVEKYLKLOINNLITNYL	60
Db	1	MSLVNRKQLEKMANVRFTQDEYVALIDALEYHNHMSSENTVEKYLKLOINSLTDIYI	60
Qy	61	NTYKSGRNKALKKFKVEYLWVLELKNLSITPVEKNLHFITWIGGOINDTAINYNQWKD	120
Db	61	DTYKSGRNKALKKFKVEYLWVLELKNNLITPVEKNLHFVWIGGOINDTAINYNQWKD	120
Qy	121	VNSDYTVKVFYDSNAFLINTLKKTIVBSATNNTLESFRENLANDPDFYKFKRMEIYY	180
Db	121	VNSDYVNAVVFYDSNAFLINTLKKTIVBSAINDTLESFRENLANDPDFYKFKRMEIYY	180
Qy	181	DQKHFIIDYKSOITBENPEFIIDNIIKTYLSNEYSKDLBALNKYIBESLNKITTANGNDI	240
Db	181	DQKQFINIYKAOBENPELIIDDIIVKTYLSNEYSKIDEIELANTYIBESLNKITTQSGNDV	240
Qy	241	RNLXFADEDLVRLYNQSELVERWNLAAASDILIRISMLKEDGGVYLDVDTLPGIQDPLFKS	300
Db	241	RNFEPFKONGSFNLYEQSELVERWNLAAASDILIRISALKEIGGYLDVDMIPGIQDPLFKS	300
Qy	301	INKPDSITNTGEMIKLEAIMKYKEYIPGYTSKNFOMLDEEVOQSFPESALSCKSDKSEIF	360
Db	301	IEKPSVVTDVFWMTKLEAIMKYKEYIPEYTSHFOMLDEEVOQSPESVLASKSDKSEIF	360
Qy	361	LPDDIKVSPLEVKIAPANNVINOALISLKDSCDVLINQIKRYKILNDNLNPSINE	420
Db	361	SSLGMEASPLEVKIAPFNKGIINQGISLVKDSYCNSLIVKQIENRYKILNNLSNPAYSE	420
Qy	421	GTDENTTKMIFSDKLASITSNEDNMFMKLTNYLKVGFAPDVSTNLNLSGPGVYTGAYQD	480
Db	421	DNDFTNTTTFIDSIMANANADGRFMWELGYLURVGFPDPVKTITNLNLSGPAYAAAYQD	480
Qy	481	LLMFKDNSTNIHLLEPELURNFPFKTKISQLTQEITSLWSFNQBARAKSQFEEYKKGYYE	540
Db	481	LLMFKEGSMIHLIEADLRNFEISKTNISQSTEQEVALSWSPDDARAKAQFEEYKKNYYE	540
Qy	541	GALGEDNDLDFRQNTVLDKOVVSKILSSWKTRNKEYIHYIVOLQDGKISYSEASONLPSK	600
Db	541	GSLGEDNDLDFSQNTVLDKYLEKISLSLARSSBERGYIHYIVOLQDGKISYSEACNLFAK	600
Qy	601	DVSSITLYQKNIEGETAYYYVADAETKETDKYRIPYQISNKRNTKLFIHGKSEFT	660
Db	601	TYDVSILFOKNIJEDSEIAYYVPGGEIQEIDKTKPSSIISDRPKIKLFIHGKDFNT	660
Qy	661	DTFANLDVDSLSEIETIINLAKADISPKYIETNLGCMNFSYSISAEETYPGKLLLKIK	720
Db	661	DIFAGFDVDSLSTEIEAAILDAKEDISPKSIEINLSCNMFSYSINVEETYPGKLLLVK	720
Qy	721	DRVSELMPSISODSTVTSANQVEYRINEBGRREILDHSGKWINKBESIIKDISSEYISF	780
Db	721	DKISIELMPSISODSIVTSANQVEYRINSEGRELLDHSGEWINKBESIIKDISSEYISF	780
Qy	781	NPKENKIIVKSKYLHELSTLQETIRNNANSDDILEKKWMLTCEINVASNIDROIVGR	840
Db	781	NPKENKITVYKKNLPBELSTLLQETIRNNSSDDIELEKVMLTCEINVASINDTQIVGR	840
Qy	841	IEEAKNLTSDSINYIKNBFKLTIESISDLYDLKHQGLDDSHIFSPEDSKTENGFRIRF	900
Db	841	IEEAKNLTSDSINYIKDFBKLTIESISDALCDLKQONELEDSHIFSPEDSETDEGSIFF	900
Qy	901	INKETGNSIFITEKEIIFSEYATHISKESINIKOTIFDNVNGKLVKKVNLDAAEVNTLN	960
Db	901	INKETGESIFVETEXTIIFSEYANHITESIKIGTIFDTVNGKLVKKVNLDTTHEVNTLN	960

Qy	961	SAFFTQSLLEYNTTKESLNLVAMKVQVYAOLFSTGLNTIITDASKVVELVGTALDETID	1020
Db	961	AAFFIQSLLEYNSKESLNLVAMKVQVYAOLFSTGLNTIITDAKKVVELVGTALDETID	1020
Qy	1021	LLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQBIEAKIGIMAVNLTAASATVIT	1080
Db	1021	LPLTSEGLPIIATIIDGVSLGAAIKELSETSDPLLQBIEAKIGIMAVNLTTATIT	1080
Qy	1081	SALGIASGFSIILLVPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETEGAFLLD	1140
Db	1081	SSGIASGFSILLVPLAGISAGIPSLVNNELVLRKATKVDFYFKHVSIVETEGVFTLLD	1140
Qy	1141	DKTIMPQDDLVTSEIDFNNNSTLKGCEIWRABGGSGHTLTDDIDHFFSPSITYYKPMWL	1200
Db	1141	DKTIMPQDDLVTSEIDFNNNSIVLKGCEIWRMEGGSGHTVTDDIDHFFSAPSITYYEPHL	1200
Qy	1201	SIYDVUNIKKEKIDFSKOLMWLPNAPNRVFGVEMGHTPGFRSLDNDGTKLLDRDRHYEG	1260
Db	1201	SIYDVLEVOKEELDLSKOLMWLPNAPNRVFAWETGWTGRLSLENDGTKLLDRDRNYEG	1260
Qy	1261	QFYWRYPATADALITKLPRYEDTNVRINLQGNTRSFIVPVTTBOIRKNTLSYSFYGSG	1320
Db	1261	EFYWRYPATADALITTLPRYEDTNIRINLQGNTRSFIVPITTBYYIEKUSYFYGSG	1320
Qy	1321	GSYSLSIPYNNMIDLNLVENDTWDVDNVVKNITTESDEIQKGBLIENILSKLNIEDN	1380
Db	1321	GTVALSLOYNNMGINTELSEVDWIIDVNVVDVTIESDKIKKGBLIEGILSTLSEEN	1380
Qy	1381	KIILNHNITFYGDINESNRFLTSLTESILEDINIIEIDLVSXKYLISGNCWKLIENS	1440
Db	1381	KIILNSHEINFSGVNGSGVSLTSLTESILEGINAIEVDLLSKSYKLLISGELKILMNS	1440
Qy	1441	SDTQOKIDHIGFNGEHOXKIPYSYINETHKYNGFIDYSKKEGLFTAFESNESIIRNIYMP	1500
Db	1441	NHIQKIDYIGFNSELQKNIPIYSFVDSGKENGFGINSTKEGLFVSELFDWLISKVYMD	1500
Qy	1501	DSNNLFTYSKDIKDJIRIINKGVOKLLIGNYFKDDMKVSLFTIEDTNTIKNGVYLDEN	1560
Db	1501	DSKPSFGYSNNLKVVKVITKDNVNILTGYYLKDDIKISLSTLQDEKTIKUNSVHLDES	1560
Qy	1561	GVAQILKFNNAKSALNTSMLNFLESINIKNIFYNNLDPNTEFIDLTNFIISGNSIG	1620
Db	1561	GVAELIKFMR-KCNTNTSDLSMLSFLESNAIKSIFVNFLOSNTKFILDANFIISGTTISG	1619
Qy	1621	QFELICDKDKNIQFYFINFKIETS YTLVYGNFSQNIIVBPSYHLDSDGNSISSTVINFSQK	1680
Db	1620	QFEFICDENDNIQFYFKFNTLETNTVTLVYGNFQNNIVERNPYDLDDSGDISSTVINFSQK	1679
Qy	1681	YLYGIDRYNKVLIAPNLYT	1700
Db	1680	YLYGIDSCVNKVVISPNIYT	1699

RESULT 2
AAW68388
ID AAW68388 standard; protein; 2366 AA.

AA
AC AAW68388;

DT 07-DEC-1998 (first entry)

DE Clostridium difficile toxin B.

Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen; pseudomembranous enterocolitis.

OS Clostridium difficile.

PN WO9808540-A1.

05-MAR-1998.

X

RESULT 2

AAW68388

ID AAW6

XX

AC AAW6

XX

DT 07-1

100

DE
CLOS

XX

KW Ant:

KW
 YY
 pser

500

OS
CLOS

XX 1700

PN YY

XX PD 05-1

XX
FD
7-50

PF 28-AUG-1997; 97WO-US015394.
 PR 28-AUG-1996; 96US-00704159.
 PA (OPHI-) OPHIDIAN PHARM INC.
 XX Williams JA, Thalley BS;
 XX N-FSDB; AAV30561.
 DR WPI; 1998-230234/20.
 DR N-FSDB; AAV30561.
 PT Host cell containing recombinant expression vector encoding Clostridium
 PT botulinum type B or E toxin - useful to treat humans and other animals at
 PT risk of intoxication with clostridial toxin.
 XX
 FS Example 18; Page 241-249; 429pp; English.
 XX
 CC This is the amino acid sequence of Clostridium difficile toxin B, deduced
 CC from the coding region (see AAV30561) of the toxin B gene. Fragments of
 CC the toxin B gene have been cloned into various prokaryotic expression
 CC systems, and assessed for the ability to express recombinant toxin B
 CC protein in E. coli. It would be advantageous to use simple and
 CC inexpensive prokaryotic expression systems to produce and purify high
 CC levels of recombinant toxin B for immunisation purposes. The invention
 CC specifically relates to recombinant proteins derived from Clostridium
 CC botulinum toxins (see AAV68389-400) and their use as immunogens for the
 CC production of vaccines and antitoxins
 XX
 SQ Sequence 2366 AA;

Query Match 77.9%; Score 6756.5; DB 2; Length 2366;
 Best Local Similarity 76.5%; Pred. No. 0;
 Matches 1301; Conservative 192; Mismatches 206; Indels 1; Gaps 1;

QY 1 MNLVYKQLOQWVKKRIODEVAILNALAEYHNMSSESVVEKYLKNDINNLTDNYL 60
 DB 1 MSLVNRQLEKMANVRFTQDEVAIDLDALEEHNMSENTEVVEKYLKNDINSLTDYI 60
 QY 61 NTKKSGRNKALKKFEYLTMVLELKNLTPVEKNLHPFWIGGQINDTAIYINQWKD 120
 DB 61 DTKKSGRNKALKKFEYLVTEVLELKNLTPVEKNLHPFWIGGQINDTAIYINQWKD 120
 QY 121 VNSDYTVKVDNSAFNLINLTKTIIVESATNTLESRENLDNPEPDYKPKRMELIY 180
 DB 121 VNSDYNVNFVDSNAFLINLTKTIVESAINDTLESRENLDNPEPDYKPKRMQIY 180
 QY 181 DKQHFHDYKSOJTEENPERIIONIKYLSNEVSKOLEALNKYIEESLNKIPANNNDI 240
 DB 181 DKQNFYNYKAQREENPELIIDIVKTYLSNEYSKEIDELNTYIEESLNKIPONSGNDV 240
 QY 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQLPKS 300
 DB 241 RNPEFNGESFNLYEQLVERWNLAASDILRISALKEGGVYLDVILPGIQLPKS 300
 QY 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLBEVORSPESALSSKSKSEIF 360
 DB 301 IEPSSVTVDWEMTKLEAIMKYKEYIPEVTSEHFDMLDEVOSSFESVLASKSKSEIF 360
 QY 361 LPDLIDIKVPLEVIAFANNSVINOALISLSDVCSDLVINOIKNRYKIILNDNINSINE 420
 DB 361 SSLGMEASPLEVIAFANNSVINOALISLSDVCSDLVINOIKNRYKIILNDNINSINE 420
 QY 421 GTDENTMTKPSDKLASINEDNMFMKITYNLKVGFPADVSTNLSGPGVYTGAYQD 480
 DB 421 DNDENTMTKPSDKLASINEDNMFMKITYNLKVGFPADVSTNLSGPGVYTGAYQD 480
 QY 481 LLMEKDNSTNHLLEPELRNPEFKTKISOLTEOITSLWSFNOARAKSOFEEYKKGYPE 540
 DB 481 LLMEKESGMNHLLEADLRNPEIKTNISQSTEGEMASLWSFDDARAKAFAPEYKKNYPE 540
 QY 541 GALGEDNDLPAQNTVLDKDYVSKKILSSMKTRNKEYIHYIVQLQGDKISYEASCNLFSS 600
 DB 541 GSLGEDNDLDFSNIVVDKYLEKISSLARSSRGYIHYIVQLQGDKISYEACNLFAK 600

QY 601 DPYSSILYQKNIEGSETAYYYVADAEIKEDKYRIPYQISNKNKIKLTFIGHGKSEBPT 660
 DB 601 TPYDSVLFQKNIEDEIAYYVADAEIKEDKYRIPYQISNKNKIKLTFIGHGKSEBPT 660
 QY 661 DTFANLDVDSLSSEIETILNLAADISPKYIEINLLGCMFSPYSISABETYPGKLLKIK 720
 DB 661 DIFAGFDVDSLSTEIEAIDLAKEDISPKSIEINLLGCMFSPYSINVEETYPGKLLKAVK 720
 QY 721 DRVSELMPSISQDSITVSAQOYEVRIINDEGKREILDHSGKWKINKEESIIKDISKEYISF 780
 DB 721 DKISELMPSISQDSIIVSAQOYEVRIINSEGRRELLDHSGEWINKKEESIIKDISKEYISF 780
 QY 781 NPXENKIIVKSKYLHELSTLLOEIRNNANSDDILEKKVLMTECEINVASNIDQIVBGR 840
 DB 781 NPXENKIIVKSKNUPELSTLLOEIRNNANSDDILEKKVLMTECEINVISNIDQIVBGR 840
 QY 841 IEAEKNLTSDSINIKNEFKLIESISDSLYDLKQNGLDDSHFISFEDISXTENGFRIRF 900
 DB 841 IEAEKNLTSDSINIKDFKLEBSISDALCDLQKQNELEDHSHFISFEDISDTGSGFSIRF 900
 QY 901 INKETGNSIFITEKEIPSEYATHISKEISNIKOTIFDNVNGKLVKKNLDAAEHVNTLN 960
 DB 901 INKETGESIFVETEKTIPEYANHITEBSIKIGTIFDTVNGKLVKKNLDTTHEVNTLN 960
 QY 961 SAFFIQSILIEYNTKESLSNLSVAMKVQVYLAQLSTGLNTITDASKVVELYSTALDETID 1020
 DB 961 AAFPIQSLEIYNSKESLSNLSVAMKVQVYLAQLSTGLNTITDAAKVVELYSTALDETID 1020
 QY 1021 LLPTLSEGLPIIATIDGVSIGAAKELSETNDPLRQEIIEAKIGIMAVNLTAATASTAVT 1080
 DB 1021 LLPTLSEGLPIIATIDGVSIGAAKELSETNDPLRQEIIEAKIGIMAVNLTAATAIT 1080
 QY 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELILOKATKVIDYFKHISLAETEGFTLLD 1140
 DB 1081 SSLGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVIDYFKHISLAETEGFTLLD 1140
 QY 1141 DKIMPODDLVLSEIDFNNSITIGKEIEMPAEGSGHTLDDIDHFFSSPSIYRKPWL 1200
 DB 1141 DKIMPODDLVLSEIDFNNSIVLGCIEIMWEGSGHTVDDIDHFFSAPSITYRBPWL 1200
 QY 1201 SIYDVLANIKKEDIPSKDLMLVLPNAPRVFGYENGWTPGFRSLDNODCTKLLDRDRHYEG 1260
 DB 1201 SIYDVLEVKEDLSDKDLMLVLPNAPRVFAWGTGTPGFRSLDNODCTKLLDRDRHYEG 1260
 QY 1261 OFYWRYPAFIADALITKLPRYEDTNVRINDGNTRGFIVPVIITEQIRKLSYSFYGSG 1320
 DB 1261 EYWRYPAFIADALITKLPRYEDTNVRINDGNTRGFIVPVIITEVIREKLSYSFYGSG 1320
 QY 1321 GSYLSLSPYNNIDNLVENDTWIDVDNVRVKNITTESDEIQGELTIENILSKNIEDN 1380
 DB 1321 GTYALSLSQYNGINIELSESDVMIIDVDNVRVDTIESDKIKKGLIEGILSTLSEEN 1380
 QY 1381 KIILNNHTNIPYGOINESRFSITFSTILEDINIIEIDLVSQYKILLSGNCKMLTENS 1440
 DB 1381 KIILNSHEINFSQGVNGSVLSTFSTILEGINAIEVDLLSKYKILLSGELKILMNS 1440
 QY 1441 SDIQOQIDHIGFNGEHQKYPYSYIDNETKNGEIDYSKKEGLTAEFSNESIIRNYMP 1500
 DB 1441 NHIQOQIDYIGFNGELQKYPYSYFVDSGKNGFNGSTKEGLFVSELPDVLISKVYMD 1500
 QY 1501 DSNNFLIYSSKDLKDIRIINKGDKVLLTIGNYFKDKMVKLSFTTIEDTNTIKNGVYLDEN 1560
 DB 1501 DSKPSFGYYSNNLKVVKITKDNVNILTYGLKDDIKLSLTLQDEKTIKLSNVHDES 1560
 QY 1561 GVAQYALKFEMNAKSAINTNSLMPLESINIKNIFYNNLDPNIBFLDTNFIISGNSIG 1620
 DB 1561 GVAELTKFMRN-KGNTNTSDSLMSFLESMNKSIFVNFVFNQSNIKFILDANFIISGTSIG 1619
 QY 1621 QFELICDKDKNIQPYFNFKIKETSITLYVGNRQNLIVEPSYHLLDDSGNISSTVINFSQK 1680
 DB 1620 QPEFICDENDNIQPYFNFKIKETSITLYVGNRQNLIVEPSYHLLDDSGNISSTVINFSQK 1679

```

QY 1681 YLYGIDRVYKVIAPNLYT 1700
Db 1680 YLYGIDSCVKNVVISFNLYT 1699

RESULT 3
ID AAR95016 standard; protein; 2710 AA.
XX AAR95016;
AC AAR95016;
XX
DT 16-OCT-2003 (revised)
DT 08-JUL-1996 (first entry)
XX
XX C. difficile toxin A.
XX
XX Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin; vaccine;
XX diarrhoea; diagnosis; therapy.
XX
XX Clostridium difficile; VPI strain 10463 (ATCC 10463).
OS
XX WO9612802-A1.
XX
XX 02-MAY-1996.
XX
XX 23-OCT-1995; 95WO-US013737.
XX
XX 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.
XX
XX (OPHI-) OPHIDIAN PHARM INC.
PA
PI Williams JA, Padiye NV, Kink JA, Thalley BS, Stafford DC;
PI Firca JR;
XX
XX WPI; 1996-230603/23.
DR N-PSDB; AAT29248.
XX
XX Fusion proteins comprising non-toxin protein and part of toxin - useful
PT to form anti-toxins against Clostridium botulinum type A, and C.
PT difficile type toxins, and to treat C. difficile intoxication, partic.
PT diarrhoea.
XX
XX Claim 63; Page 290-302; 434pp; English.
XX
XX Clostridium difficile VPI strain 10463 toxin A (AAR95016), the product of
CC the toxin A gene (AAT29248), is a potent cytotoxin that plays a direct
CC role in damaging gastrointestinal tissues and is associated with
CC diarrhoeic disease. It can be obcd. by expression in transformed E. coli
CC hosts of portions of DNA that together cover the entire toxin A gene.
CC Toxin A, and portions of it (see also AAR95014-15 and AAR95017), pref.
CC expressed as fusions to polylhistidine affinity tags or maltose binding
CC protein, are used to raise avian antibodies useful as antitoxins or
CC diagnostics, and in vaccine prodn. (Updated on 16-OCT-2003 to standardise
CC OS field)
XX
XX Sequence 2710 AA;

Query Match 51.5%; Score 4467; DB 2; Length 2710;
Best Local Similarity 50.7%; Pred. No. 5e-225;
Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;

QY 1 MNLVNAQLOKMYKFRQDEYVAILNALSEYNNMSESSVVEKYLKLOINNLTNYL 60
Db 1 MSLISKEELIKLAY-SIRPRENEYKITLNLDEYNKLTNNENKYLQLKXLESIDVFM 59
XX
QY 61 NTYKSGRNKALKKFKYELTMEVLELKNLSLTPVEKNLHFIWGGQINDTAINYNOWKD 120
Db 60 NKYKTSRNRALSNLKDLKVELIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWAD 119
QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNTNLTLESFRENLDNDFDYNKFKRMEIY 180

```


Db 1198 HIPSLSYSAIGIETENLDFSKKIMPLNAPSRVFWWETGAVPGLRSLNDGTRLLDSYR 1257
 Qy 1256 DHEGQYRYFAPFADALTKUKPRYEDVNRINLDGNTSRFVIVITEQIRKNSYS 1315
 Db 1258 DLYPGKPYRFYAFYF-DYAITLKPVEDTNKIKLDKOTRNPIMPTITTTNEIRKNSYS 1316
 Qy 1316 FYSGGYSYSLSPYNNIDNLAVENDTWIDVDNVVKNITTESDEIQGELIENTLSKL 1375
 Db 1317 FDGAGGTYSLLSSYPSTINUSKODLWIFNIDNEVREISBNGTIKKGKJIKOVLSKI 1376
 Qy 1376 NIENKIILANNHTINFGDINESRFLSFLSILEDINIIEIDLVSYSKYILLGNCKMK 1435
 Db 1377 DINKKLIIGNOTDFSDIDNDRYIFLTCELDDKSLIIEINLVAKSYSLSSGDKY 1436
 Qy 1436 LIENSIDQKIDHIGFNGHQKYPYSYID-NETKYNQFIDYSKXEGFLTAEFNSHII 1494
 Db 1437 LISNLNTIEKINTGLD---SKNIANYTDESNNKYFGAI-----SKTSQKSII 1483
 Qy 1495 RNIYMPSSNNI-----FYSSKDL--KDIRIINKGDKLIGNYFKD---DMKVSLS 1541
 Db 1484 H--YKDSKNILEFYNDSTLEFNSKOFIABEDINVMFKDDINTITGKYVDNNTDKSIDFS 1541
 Qy 1542 FTIEDTNIKLNGVYLDENGVAQILKPMNNAKSAALNTSLSMNFLESINIKFIYNNLDP 1601
 Db 1542 ISLVSKQVKNVGYLNEVSYSVYLDYFKNSDGHNTSNFNLFLDNISFWKLPGE--- 1598
 Qy 1602 NIEBILDTNFTIISGNSIGOFELICDKDKNIQPIYFNFKIKETSYTLYGVNQNQLIVEPS 1661
 Db 1599 NINFDYKFLVGTNLNGYVEFICDNKNKIDIVFGEMKTSKSTIFSNGRNVVVEPI 1658
 Qy 1662 YHLDOSNISSTVNFQKLYGIDRVVKNVVIAPNLYT 1700
 Db 1659 YN-PDTGEDISTLDFSEPLYGIDRYINKVLIAPDLYT 1696

RESULT 4

AAW68387
 ID AAW68387 standard; protein; 2710 AA.
 XX
 AC AAW68387;
 XX
 DT 07-DEC-1998 (first entry)
 XX
 DE Clostridium difficile toxin A.
 XX
 KW Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen;
 KW pseudomembranous enterocolitis.
 XX
 OS Clostridium difficile.
 XX
 PN WO9808540-A1.
 XX
 PD 05-MAR-1998.
 XX
 PF 28-AUG-1997; 97WO-US015394.
 XX
 PR 28-AUG-1996; 96US-00704159.
 XX
 PA (OPHI-) OPHIDIAN PHARM INC.
 XX
 PI Williams JA, Thalley BS;
 XX
 DR WPI; 1998-230234/20.
 DR N-PSDB; AAV30560.
 XX

Host cell containing recombinant expression vector encoding Clostridium botulinum type B or E toxin - useful to treat humans and other animals at risk of intoxication with clostridial toxin.

Example 15; Page 220-230; 428pp; English.

This is the amino acid sequence of Clostridium difficile toxin A, deduced from the coding region (see AAV30560) of the toxin A gene. Toxin A is a

CC potent cytotoxin that plays a direct role in damaging gastrointestinal
 CC tissues. Severe cases of C. difficile intoxication result in
 CC pseudomembranous colitis. This would be prevented by neutralising the
 CC effects of toxin A in the gastrointestinal tract. Examples are provided
 CC of the production of recombinant C. difficile toxin A in host cells and
 CC of the in vivo neutralisation of toxin A by antibodies against
 CC recombinant toxin A polypeptides. The invention specifically relates to
 CC recombinant proteins derived from Clostridium botulinum toxins (see
 CC AAW68389-400) and their use as immunogens for the production of vaccines
 CC and antitoxins
 XX
 SQ Sequence 2710 AA;
 Query Match 51.5%; Score 4467; DB 2; Length 2710;
 Best Local Similarity 50.7%; Pred. No. 5e-225;
 Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;
 Qy 1 MNLVNAQLOQWVYVKFRIQDEYVAIINALBEEYHNNESSVWEKYLKLDKNNLTNYL 60
 Db 1 MSLISKEELIKLAY-SIRPRENEYKITLNDENKLTNNENKYLQKLNESIDVFM 59
 Qy 61 NTYKSGENKALKKFKYLTMEVLELKNNSLTPVEKNLHFWIGGQINDTAINYNQWD 120
 Db 60 NKYKTSNRNALNKDILAEVILKNSNTSPVEKNLHFWIGGEVSDIALEYIKQWAD 119
 Qy 121 VNSDVTYVYDSNAPLINTLKKIVESATNNTLESFRENLDNDEPDKYKRGWEIYY 180
 Db 120 INAEYNIKLYDSEAFVLNLTAKAIVESSTTEALQLLEEIEQNPFQNNKFKYKGEFIY 179
 Qy 181 DKQHFIDYKSOEENPEFIIDNIIKTYLSNEYSKOLEALNKYIEESLNKITANNNDI 240
 Db 180 DRQKFIINYKQINKPVTPTIDDIKSHLVSEYRNDETVLESYRTNLRKINSVHGDI 239
 Qy 241 RNLEXFADBDLVRLYNQELVERWNLAAASDIIRISMLKXEDGGVLDVLDLPQIDLPKS 300
 Db 240 RANSLFTEQELLNYSQELLNAGNLAASDIVRLALAKNFGGVLDVLDMLFGIHSDLFKT 299
 Qy 301 INKPSDITNTGEMIKLEAIMKYEYIPGYTSKNFMDLDEEVQKSFESALSKSKSEIF 360
 Db 300 ISRPSSIGLDREWEMIKLEAIMKYKYINNYTSNPFKLDQQLKDNFKLIESKSEKSEIF 359
 Qy 361 LPDDIKVSPLEVKIAPANNVINOALISLKDSYCSDLVINQIKRYKILNDNLPSINE 420
 Db 360 SKLENLVSDLEIKIAFALGVSINOALISKOGSYLTNLVIEQVKRYOFLNQHLPALAES 419
 Qy 421 GTDENTTKMIFSDKLASISNEDNMFMKITNYLKVGFAPDVRSTINISGPOVYTGAYD 480
 Db 420 DNNFTDTTKIFHDSLFNGATAENSMLTKIAPYLOVGFMPPEARSTISLSGQAYASAYD 479
 Qy 481 LLMFKDNSTNIHLLEPELRNFEFPFKTKISQLTQEITSLWSFNQARAKSQPEYKKGYPF 540
 Db 480 FINLQENTIEKTLKASDLIEFKFPENNLSQLTQEINLSWSPDQASAKYQEKYVRYDYG 539
 Qy 541 GALGEDDNLDPQNTVLDKDY-VSKILSS--MKTRNKEYIHYIVOLQDKITSYASCNL 597
 Db 540 GSI SEDNGVDYFNKNTALDKVLLNNKIPSNNVVEAGSKNYHYI1QLQDDI-SYEATCNL 599
 Qy 598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKIDKYRIPQISNKENIKLTFIGHCK 655
 Db 600 FSKNPKNSIIIQNNM--NESAKSYFLSDGSEILELNKYRIPERLNKKEKVVTFIGHCK 657
 Qy 656 SEFNTDTFANLDVDSLSSESIETILNAKADISPKYIEINLLGCNMFYSISAEETYPGKL 715
 Db 658 DEFNTSEFARLSDVLSNEISSFLDTIILKDISPKNVEVNLGCNMFSDVFNVEETYPGKL 717
 Qy 716 LKIKDRVSELMFSPISQDSITVSANQYEVREINEGKREILDHSGKWINKESIIKIDSK 775
 Db 718 LLISMDKITSTLPDVNKNISITIGANQYEVREINSEGRKELLAHSGKWINKKEAIMSDLSK 777
 Qy 776 EYISNPKENKIIIVKSKYLHELSLLOEIRNNANSDDILEKVMMLTECEINVAINDRQ 835
 Db 778 EYIFFSDIDNKLAKSKNIPGLASISEDIKITLLDASVSPDTKFINLNLKLIENSIIGDY 837

836 IVEGRIBAKNLTSDSINVIRKNEFKLIESISDSLYDLKHQGLDSDHFIQFEDISKSTENG 895
 838 IYKLEFPVKIHIHNSIDDLDEFNLEWDEIYELKLNLDKYLIFEDISKUNST 897
 896 FRIRPINKETGNSIFETKEIFSBYATHISKEISNIKOTIFDNNVNGKLVKKNVLADE 955
 898 YSRPINKSNGESVYVETKEIFSKYSEHITKEISTIKNSIITDVGNNLLDNIQLDHTS 957
 956 UNTLSAFPIQSLIYNTTKSLSNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTAL 1015
 958 VNTLNAAFPIQSLIDYSSKNVLDLSVQVQYQALFSTGLNTIYDSIQLVNLISNAV 1017
 1016 DETIDLLPLSGPLIATIDGVSGLGAAIKELSETDPLLRQIEAKIGIMAVNLTAAS 1075
 1018 NDTINVLPTIEGIFVITLDGILGAAIKELDEHDPDLKKELEAKVGVLAINMLSI 1077
 1076 TAVTSALGIAGSFGFILLVPLAGISAGIPSLVNNELILODKATKVIDYFHRHISLAETGA 1135
 1078 RATVASIVGIGAEVIFELPLIAGISAGIPSLVNNELILODKATSVVNFHLSKSKYGP 1137
 1136 FTLLDOKIIMPDDILVSEIDFNNSITLGCIEIWRAGGSHLTDIDHFFSPSPITY 1195
 1138 LKTEDDKILVPIDDILVSEIDFNNSIKLGTGILNLAEGSGGHTVTGIDHFFSPSSISS 1197
 1196 RKPWLSIYDVLNKKKEKIDFSDKDLAVLPNAPRVFEGYMGWTPGFRSLDNDGTLLDRIR 1255
 1198 HIPSLSIYSAIGIETENLDFSKIMPLNAPSFPWETGAVPGULRSLENDGTLLDSIR 1257
 1256 DYVEQFVWRPAFTADALITKLPREYEDTVIRNLGNTSFIQVITTEQIBKRLSYS 1315
 1258 DLYPKFVWRVAPP-DYAITTLKPYEDTNIKIKLDKTNFIMPTITTEIRNKLSYS 1316
 1316 FYGSGSYLSLSPYNNMIDNLVENDTWIDVNVNKNITIESEIOKGBELIENILSKL 1375
 1317 FDGAGTYSLSSYPISITNLSKDLWIPNIDNEVEISIENTIKKKLIKDLVLSKI 1376
 1376 NIEDNKILNNTIPIYGDINESFISLTSLEIDNIIIEIDLVSYSKILSGCMK 1435
 1377 DINKKLIIGNQITDFSGDIDNKRIYIFLTCELDKISLIEINLVAKSYSLLSGDKNY 1436
 1436 LIENSDDIOQKIDHGFENGHOKYIPYSYID-NETKNGFTDYSKKEGLFTAEPSENGII 1494
 1437 LISNUNTEKINTGLD---SKNAYNTDESNNKYFGAI-----SKTSQSKII 1493
 1495 RNIMPDSNNL-----FIYSSKDL--KDIRIINKGVDKLLIGNYFKD---DMKVSLS 1541
 1484 H--YKDSKNILEFYNDSTLENSKDFIAEDINVPKDDINTITGKYVDNNTDKSIDFS 1541
 1542 FTIEDNTIKNGVLDENGVAQILKFMNNAKALNTSNLSNFIENIKNIIFVNNLDP 1601
 1542 ISLVSKNQVKNGLVNESVYSYLDLVKNSDGHNTSNFMNLFILDNISFWKLFGFE--- 1598
 1602 NIEFLDTNFIISGNSIQPFLICDKDKNIQPIYFINKETSITLVGNRONLIVEPS 1661
 1599 NINFVIDKFTLVGKTNLGYVEFICDNNKNIDIFYGEWKTSSKSTIPSGRNVNVEPI 1658
 1662 YHLDGSGNTSSVINFPSQVLYGIDRYVNNKVIAPNLYT 1700
 1659 YN-PDGEDISTSLDPSFPLGIDRYINKVLIAPDLYT 1696

RESULT 5

AA33700 standard; protein; 546 AA.

XX AA33700;

XX 18-JAN-2000 (first entry)

XX C. sordellii lethal toxin protein fragment.

XX Lethal toxin; immunotoxin; antitumor; glucosyl transferase; glycosylate;
 KW GTPase; Ras; epidermal growth factor stimulated MAP-kinase; inhibition;

KW signalling pathway; cell-specific toxin; treatment; cancer.
 XX Clostridium sordellii.

XX DE19802569-A1.

XX 09-SEP-1999.

XX 23-JAN-1998; 98DE-01002569.

XX 23-JAN-1998; 98DE-01002569.

XX (UYPR-) UNIV FREIBURG ALBERT-LUDWIGS.

XX Aktories K, Hofmann F;

XX WPI; 1999-509323/43.
 XX N-PSDB; AA223800.

XX New fragment of the lethal toxin from Clostridium bacterium, useful for
 PT treating cancer.

XX Claim 1; Page 7-9; 14pp; German.

XX This invention describes a novel fragment (I) of the lethal toxin (IT) of
 CC Clostridium sordellii which has antitumor activity. (I) is a glucosyl
 CC transferase that glycosylates, and thus inactivates, GTP(guanine
 CC triphosphate)/ases, particularly Ras (an oncogenic product overexpressed
 CC in many tumors), resulting in inhibition of epidermal growth factor
 CC stimulated MAP-kinase signalling pathways. (I), particularly in the form
 CC of immunotoxins, are used as cell-specific toxins, particularly for
 CC treating cancer. When included in immunotoxins, (I) can be targeted to
 CC selected cells. Compared with the complete IT, (I) is smaller, so enters
 CC cells more easily, resulting in greater toxicity in the cytosol, is less
 CC likely to induce formation of (neutralizing) antibodies, and is more
 CC active than the holotoxin. This sequence represents the lethal toxin
 CC fragment described in the invention

XX SQ Sequence 546 AA;

Query Match 32.2%; Score 2792; DB 2; Length 546;
 Best Local Similarity 99.5%; Pred. No. 3-7e-138;
 Matches 543; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MNLVNAQQLQKVVYKFRIOEDYVAILNALBEYHNMSSESVVEKYKLDNNLTNDYL 60
 Db 1 MNLVNAQQLQKVVYKFRIOEDYVAILNALBEYHNMSSESVVEKYKLDNNLTNDYL 60
 Qy 61 NTYKSGRNKALKKPKVEYLTMEVLELKNNSLTPVEKNLHFIIWGGQINDTAINYINQWD 120
 Db 61 NTYKSGRNKALKKPKVEYLTMEVLELKNNSLTPVEKNLHFIIWGGQINDTAINYINQWD 120
 Qy 121 VNSDVTVKYFDSNAFLINTLTKTIVESATNNTLESFRENLDNPEFDYKFKRMEIY 180
 Db 121 VNSDVTVKYFDSNAFLINTLTKTIVESATNNTLESFRENLDNPEFDYKFKRMEIY 180
 Qy 181 DKQKHFDYKQIEENPEFIIDNIITKTYLSNEYSKDLKALNKYIEESLNKITANNNDI 240
 Db 181 DKQKHFDYKQIEENPEFIIDNIITKTYLSNEYSKDLKALNKYIEESLNKITANNNDI 240
 Qy 241 RNLEKFADEDLVRLYNQELVERNWLAAASDIILRISMLKEDGGVYLDVDILPGIQPLFKS 300
 Db 241 RNLEKFADEDLVRLYNQELVERNWLAAASDIILRISMLKEDGGVYLDVDILPGIQPLFKS 300
 Qy 301 INKPSITNTSWEMIKLEAIMKYKEIPGYTSKNFMDLDEEVSQSPESALSKSDKSEIF 360
 Db 301 INKPSITNTSWEMIKLEAIMKYKEIPGYTSKNFMDLDEEVSQSPESALSKSDKSEIF 360
 Qy 361 LPJDDIKVSPLEVKIAPANNVINOALISLKDSYCSDLVINQIKRYKILNDNLNPSINE 420
 Db 361 LPJDDIKVSPLEVKIAPANNVINOALISLKDSYCSDLVINQIKRYKILNDNLNPSINE 420
 Qy 421 GTDFNTMTKIFSDKLASISNEDNMFMKITYLNKYGVGFAPDVRSTINLSGPGVYTGAYQD 480

Db	421	GTDFNTTKIFSDKLASISNEDNNMFKIITNYLKNCFAPDVRSTINLSGPGVYTGAQD	480
Qy	481	LLMFKDNSTNIHLLEPELRNFEPFKTKISQTEQBITSLMSFNQARAKSQFEYVKKGYPE	540
Db	481	LLMFKDNSTNIHLLEPELRNFEPFKTKISQTEQBITSLMSFNQARAKSQFEYVKKGYPE	540
Qy	541	GALGED 546	
Db	541	GALGED 546	

RESULT 6	
ABO14444	
ID	ABO14444 standard; protein; 3169 AA.
XX	
AC	ABO14444;
XX	
DT	23-OCT-2003 (revised)
DT	21-AUG-2003 (first entry)
XX	
DE	E. coli O157 clostridial cytotoxin-like protein o3169.
XX	
KW	O21D; acute haemorrhagic colitis; haemolytic uraemic syndrome;
KW	food poisoning; clostridial toxin-like protein.
XX	
OS	Escherichia coli; strain O157:H7.
XX	
FN	US2003023075-A1.
XX	
PD	30-JAN-2003.
XX	
PF	01-APR-2002; 2002US-00114170.
XX	
PR	04-DEC-1998; 98US-0110955P.
PR	03-DEC-1999; 99US-00453702.
XX	
PA	(BLAT/) BLATTNER F R.
PA	(BURL/) BURLAND V D.
PA	(PERN/) PERNA N T.
PA	(PLUN/) PLUNKETT G.
PA	(WELC/) WELCH R.
XX	
PI	Blattner FR, Burland VD, Perna NT, Plunkett G, Welch R;
XX	
DR	WPI; 2003-479497/45.
DR	N-PSDB; ACD19243.
XX	
PT	New DNA sequences from Escherichia coli strain O157:H7, useful for
PT	detecting E. coli O157:H7 in a sample, or in designing diagnostic probes
PT	which can be used to distinguish strain O157:H7 from strain K12 using
PT	molecular techniques.

[illegible]

Db 991 IITNAY---IKERDKL-----SGIIYNIENSIISGHESFDIIRSNQHEWGDLSSTVEQPK 1042
Qy 784 ENKIVISKYHELSLLOETRRNANSDDLEKKWMLTECEINVAS-----NIDQIIVE 838
Db 1043 KPEFYVXSE-LSSAKSFDIDKKN-YITDPETKRNVLHQDSDIKERIAFLDISHYAYP 1100
Qy 839 GRIBEAKNLTSDSINYIKNEFKLETESI-----SDSLYDLKHONGL 878
Db 1101 GSLLEKQLS-----GYVPSDINIABVLLASGVSHYGVVPAPSDKLLLELRHTK 1156
Qy 879 DSHFISPEDISKTEGFRIRINKETGNSI---FIETEKEIFSEYATHISKEISNIKOT 935
Db 1157 SNSEWI--EKITP---YVYDILSDNVSNVLRPPLSEQKKILNDIKLEISKVS---EQ 1207
Qy 936 IPDNVNGKLVKKNLDAHEVNTNSAFFQOSLIEYNTK-----ESLSNLSV----- 983
Db 1208 YFMKLTQKSSVIGIKYSVDPDRYNENLFLSLPINQNLTLFFMYRYPEMLYDIHIGIEN 1267
Qy 984 -AMKVQVYAQLFSTGLNTITDASKVVELV-----STAL-DEFIDL 1021
Db 1268 KANREFIYKFSNLDPLINDERVLNLEGLIKKYKLSUSEIHRTLTNTSTFADISPL 1327
Qy 1022 LPTLSEGLPIITATII-----DQVSLGAIAK-----ELSETNPLLR 1057
Db 1328 LQTI---CPSITITIKTEYVYGHQLTNMTVASVVKYDPSNGLAINSIDSVDYPAH 1384
Qy 1058 QEIE-AKTGIMAVNLTAATAIVTSAL-----GIASGSFILL 1093
Db 1385 TIVEQAKYNLLSWNDFYNTHASIWDTTARQKSTNIEFHPQSLFLDRDSKGKGLGLSLLY 1444
Qy 1094 VPLAGISAGISLVNN-----ELIQDKATKVIDYFKHLSLAETEGATLLDDKLIIMPQDD 1149
Db 1445 LDTGGYGGYQKLRHNDITASTLYQTK-----YNDNLKLSNRDDFFLRKTYQRIITWSNE 1498
Qy 1150 LVLSEIDFNNSITLKGKEIWAEGSGHTLTDI--DHPFSSSIYRKPWLSIYDVLN 1207
Db 1499 L-----GNRLKNAQLEVLK-----DPLTEGILYQBRISLLIITEVHSLAQOISS 1348
Qy 1208 IKK-----EKIDFKDLMVLENAPNAPRVGYEMGTWPGFRSLDNDGTKLDD 1252
Db 1549 FWRVTDNPGHCDHFSLAQALTFIKVI-----TSNR-----NFSLYSGSI--- 1589
Qy 1253 RIRHY-BQFYRWYFAFIADALTKLAPRYEDNVRINLDG---NTRSFIVPVITTEQI 1308
Db 1590 -VKIYFSESLANWKYIKLPLVQTSLLRDLYLTPEKLSLSTGGSLNMGHLVPY----- 1642
Qy 1309 RKNLSYSP-YSGSGYSLSLSPYNNIDNLVENDTWIDVNVKNITIESDIOQ--- 1363
Db 1643 -----SFIYDIGV-----INGNRISST---DVKNKIRSLKINGDILQHYIN 1682
Qy 1364 -----KGELTENILSKLNTEDNKI-----ILN---NHTINFY- 1392
Db 1683 THYLSBETQTKIIVDFLQDNTIKVKLESIDKPISEIQOPLHSILSRQEHVKLLS 1742
Qy 1393 GDINE-SNRF-----ISLTFSLIEDINI-----IIEI---DLVSKSYKILLSGNCKWL 1436
Db 1743 GLLDEFSNKLKQGLSLKTNVLSVNNFESKINSDEVTBTVDLQGLRYRVDITRVIGL 1802
Qy 1437 I---ENSSDIQKIDHIGENGHQKIPYSYDNETKNGFTDYK-----KEGLTAFESN 1490
Db 1803 TPEKINSLSALEHMN-----IDAIMSVIGLVQYARMKQNDNISAIHDHAG 1849
Qy 1491 E-SIIRNIYMPDSNLFYSSKDLADIRINKGVK-LLIGNYFKDDMKVLSFTIEDTN 1548
Db 1850 AVSDIKNV-----DKFLIGLITLNNRVNPGVGSGASLEGFTISSGLEVCAS----- 1897
Qy 1549 TIKLAGV---YLDENGVAQILKFMNNAKSALNTSNLNFLESINIKINFYNNLD----- 1600
Db 1898 -RMGCTAGRYL--SNVAKVIL-----PLLDI--GINIWSLYDSSLNHAHAKAT 1939
Qy 1601 PNIETI---LDTNF-IISGSNIGQFE-----LICDKKNIOPIYFNFKIKETSYLVY-- 1650
Db 1940 TQIEVISTAIDVSFSSINTALSIGAIVAPPLAIA-----IVPITI-PSHEVKNYAVYVQ 1993

Qy 1651 -GNRQNLIVPSYHLDD-----SGNISSTVINFSQKLYG---IDRVNKKVII 1694
Db 1994 INERHKIWEAKYLDNGSAKVLINRATGIDLSNNQVLGNLYLDRENPIL 2047

RESULT 7

ABU48941
ID ABU48941 standard; protein; 4588 AA.
XX
AC ABU48941;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #34468.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Ureaplasma urealyticum.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX N-PSDB; ACA52811.

XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 76865; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 5213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

Sequence 4688 AA;

Query Match 5.0%; Score 431.5; DB 6; Length 4688;
Best Local Similarity 19.1%; Pred. No. 3.2e-13;
Matches 420; Conservative 329; Mismatches 743; Indels 709; Gaps 103;

QY	13	VTVKPRIOBDEVAILNALE--EYNMGESSVVEKYKLKDINNLDTNYLNT---YKGGGR	68
Db	403	LYEVIQIKKTNEVDVWNLKQIPYENRS-----INNLASNALNTPYQTKNGD	449
QY	69	NKALKKFKEYLTME-----VLELKNS-----LTPVEKNLHFIWGGQINDTAIYNQWKD	120
Db	450	INLIAKAVPYVYNQOVYGIFDQNNQHQILAKYKDGDTAFDGTGALNNNSYSLDKIVS	509
QY	121	VNSDVTWVDFVDSNAFLINTLKTIIVESATNNTLESFR-----ENL-----	161
Db	510	VSNPQNVLV---SNFDLTSKQKQILKPPAANASVDSTKTKQILENLNDLNLNOKLVATFV	566
QY	162	--NDPEFYNKPYRKMIBIYDKQ---KHFIDYKYSQIBEN-----	197
Db	567	DNNDKEYKVAVNDQNNKIIFDSNDLPKGYI-VHLAKVENNDLKNVINLNDPELXDKIID	625
QY	198	-----PEPIIDN-----IITYILSNEYSKDL--EALNKY-----IEBSLN	230
Db	626	KRDLNLLSDHPDFTYDNDGNLIEHTQLANDLNDLQKALNNANVKGIVVDQDGTIEHID	685
QY	231	KITANNGNDIRNLEKFADEDLVR--LYN-QELVERNLAAASDILRI\$MLKEDGGVYLDV	287
Db	686	V\$IDANGKVILPTKLNANDPTKNIITLKVVLVKQNNQPNIDILSEELSGDNIHSPK--	744
QY	288	DILPGTQDPLFK\$INKPD-----SITNT\$WEMIKLEALMYKEYIPGYTSKNFMDLDEBQV	343
Db	745	-----KPTITAKTENDDYEISFNSPLANKIKLTFKTD-----NNTNTKTV\$AS	791
QY	344	R\$FESAL\$K\$K\$K\$E\$IFLP-----LDDIKV-----SPL-----	371
Db	792	IGLQ\$KAIFKTSDDAIFAPDHKYLTKIEADNKKVANI\$EISPLDRI\$V\$NKG\$N\$V\$NADN	851
QY	372	--\$VKIAFANNSVI-----NQALISIKDSYCSOLVINQIKRY-----KI--	409
Db	852	KHEFKIPDQKNKDLTAVYKDKNNREIHPVITKDDGKGVIVNPNNLFDENKIDYDFDKIVD	911
QY	410	LNDNLNPSINEGTDENTMKIFSD-----KLASISN--EDNMNMKIXITNLYKVG	457
Db	912	LNEYPNKTLDRNSINKDVSAINDGVQ\$KLVKAPAVSNVTVNAINFQVWLF\$DNLKLS	971
QY	458	----FAPDVRS\$INLSG\$GVVTGAYQ-----D\$LLMPKDN	487
Db	972	Y\$N\$Q\$FALITAKVNDLNDTQKIIAYDPKTN\$YKLNFDTHLDANTKYKVVDV\$ELLE\$ANK	1031
QY	488	STNIHLLEPEL\$NFP\$KTKISQUT\$E\$BIT\$LSM\$F\$N\$Q\$AR\$KQ-----F\$EYKKG\$FEG	541
Db	1032	EXPKLIKLDVLNFP\$FT\$S\$ATTIN---P\$PIWTK\$P\$DV\$V\$K\$TNDTIT\$F\$EIDDK--DN	1084
QY	542	ALGEDNLD\$PAQNTVLDKDY\$YSKIL\$SMKTRNKEYIHYI\$V\$Q\$G\$KISYE\$ASCNL\$F\$KD	601
Db	1085	ILKNDQKI-YAQLALMD\$D\$D\$D\$D\$VINPLV\$NFSKT-----NKIASINGL\$DK\$NS	1133
QY	602	PYSSILYQKNIS\$E\$TAYVYV\$ADA\$EK-----EIDKVRIPQIS\$K\$ENIKLTI\$G\$K	655
Db	1134	KY\$I-----KNL-----Y\$YN\$Q\$N\$Q\$V\$YLF\$RNDVTKYB-QH\$ITNPHKINL\$F--NK	1177
QY	656	\$E\$FNTD\$TF---ANL-----D\$V\$LS\$E\$E\$ITILNLA\$K\$ISPRY	690
Db	1178	\$AVEQ\$DIFADHANL\$FIDYKDYDQKLRINEDVKIYQ\$NIDNTK\$ELO--IGYK\$V-VANNK	1234
QY	691	IEINL\$G\$CN\$F\$YSISAE\$ETPGKLLKIKDRV\$ELMPS-----ISQ\$STIV	737

Db 1218 -----EHPATAAATRLBEAAQKANKNDPVAUKLYQ-LQOARTSSD----- 1259
QY 367 KVSPLVVKIAPANNVINQALISIKDSYCSDLVINQIKRYKIILNDNLNFSINEGDTFT 426
Db 1260 -----TALIDQIKQL-----NLQDYVYRAQYD---PILREASSVEK 1293
QY 427 T-----MKIFSDKLASISNEQNMFMKIXNYLVKGAPDVRSTINLSGPGVYTGAYQDL 481
Db 1294 TWLMOQIKAKGSAVAVQIDTALQGBERTSSGVERAASEVYORGACGTG----- 1344
QY 482 LMFKNSTNIHLLEPELANFPFKTKISQLTQEITSLWSPNQARAKSQEEVKKGYFEG 541
Db 1345 --FNDG-----LNFR--EIKIPGVD-----KMSVLEMKGRLEG 1376
QY 542 ALGEDNDLDPQNTVLDKDYVSKILSMKTRNKEYIHVQLQDQKISYEASCNLFPSK 601
Db 1377 KLTHS-----QREALSARI---AETISQAEYI-----DKVLRQTA--MPSD 1412
QY 602 FY--SSIL-----YQNIQEGSETAYYYVADAEIKEDKVRIPYQISNKNKILFTIG 652
Db 1413 FHDAGSVNRLVPQDFYLSLVGDRGTGRCYPL-----VRTWAVALAG 1454
QY 653 HCKSEFNTDTFANLDVDSLSSEIETILNAKADISPKYIEINLLGCMFS--YGISAEET 710
Db 1455 GGEAGINS-----LVQKLPASADPQAGSSTLLKNSLISLHYNVEAVQA 1498
QY 711 YPKLLKIKIDRVSELMPSISQDSITVSANQVEVRINE---EGKRE-----ILD 756
Db 1499 STELGQPNLSEVVSLAATTGTSFALNTQNHMMVGSTVTAEGERYVYFDPNVGIFAFD 1558
QY 757 HSGKWK--EBSIKDISKEYISFNPK-----NKILVSKYLHESTLLQETRNANS 810
Db 1559 NT-KSLSRAMEQLVVRGLAAHYGFGSKSVPAFNLEIDTRKMAEV-----PVGNGLV 1612
QY 811 SDIDLEKVMTECEINVASND--RQI-----VEGRIEBAKNLTSDSINYIKNEFKLIE 863
Db 1613 AD-----LTFRE-ELASVIGQRQVEQVVSQVQWYMEDLRSLPALKTFDAECWGARF 1663
QY 864 STSDSLYDLKHQGLDGHFI-----SFEDISKTEGFRIRFINKE-----TGNISF 910
Db 1664 EVASARLAQEHQL---DSRWLPIATTD--QGEGRYAVQFINRDQPEQTRWLITSSTF 1718
QY 911 IETEKEIFSEYAT--HISKEISNKTIDFNVNGKLVKKNLDAAEVNTLNSAFFIQS 967
Db 1719 VBFRFRIDEMAVLNEHFTLE-----HGRMLRGVGEAAPVDGLNAGFAVQT 1766
QY 968 LIENYTKES-----LSNLSVAKQVYAQLFSTGLNTITDASKVELVSTAL----- 1015
Db 1767 LIQWFADKNRHDAAAGVIFDLPATALKVHSYLNVLQMAHGGVQDIAKTALVTRLRGEV 1826
QY 1016 -----DETIDLLPTLSEGLPIATIIDGVSLGAAIKELSETNDPILROEIEAKIGIM 1067
Db 1827 VAAETSLKDPASNLGHTVNEGA-----GVLFQGMVGLDAYE--LAHAENDVQKAVF 1876
QY 1068 AVNLTAASTAVITSALGIASGF-----SIL---LVPLAGISAGIPSLVNNELILODKA 1117
Db 1877 GTCFLAPDSASFYTGAGVAGLGVAGTAGAVLGASVILGLGAVFTALAQAFGAEDA 1936
QY 1118 TKVIDYFKHISLAETEGATLLD-DKINMPQDLVLSEIDFNNSNITGKCEIWRAGEGS 1176
Db 1937 KAVGRYFDVDRKAYKNGVRYDNKNKVLPLAGAVIKSLDKRKHQISFDSQVIVRTHSGS 1996
QY 1177 GHTLTDIDHFF---SSPITY-RKPWLSIYDVLNKEKIDFSKDL-----MVLNPA 1225
Db 1997 --TGSGEINFFWVGDFPMVHQRQAIEVRSGIGYK-----DISPLHESQSNVILPGT 2050
QY 1226 PNRVFGYMGWPGFRLSDNGTKLLDRDRHVEGYVRYFAFTADALITLKPVEYDT 1285
Db 2051 PKSIYEWMLPGATRHADGDFVIRLLEE--DKRFDYDFVIFGEEETIRIHEHYDT 2108
QY 1286 NVRLINDGNTRGPIVPIITTEQIRKNLSYFYGSGGYSLSLSPYNNMIDLNVENDT-- 1343

Db 2109 PIEVLDQRNQLVVPPELPE--LHSFLRYEIKGAGGEYLIGL---NEGTQVKLTSDSASA 2164
QY 1344 ---WVIDVDN-----VVKNTITIESDEIQGELIENILSKLINIEDNKIILNNH 1387
Db 2165 LSRWIIDSSQIASDSISVKNQLVVGVVVELDPQTQGVVL----- 2205
QY 1388 TINFYCDINESNRFISLTFISLEIDINIIEIDLVSYSKILLSGNCKMLKINSDDIOOKI 1447
Db 2206 VVNGKGEVREVD-FAGLTQVVSE-----DAGKQVPG--QQIEQHLSDLAK-- 2249
QY 1448 DHIGNGEHQKYPISYIDNETKYN-----FDYSKKEGLFT 1485
Db 2250 -----AQLHGVVVENYS-HNGRDVGAFYDVGKERMLFT 2285
RESULT 10
ABM66965
ID ABM66965 standard; protein; 2410 AA.
XX
AC ABM66965;
DT
XX 20-NOV-2003 (first entry)
XX Photorhabdus luminescens protein sequence #62.
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
whoooping cough.
XX Photorhabdus luminescens.
XX W0200294867-A2.
PN 28-NOV-2002.
XX
XX 07-FEB-2002; 2002WO-1B003040.
XX
XX 07-FEB-2001; 2001FR-00001659.
PR (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
XX WPI; 2003-148459/14.
XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX Claim 2; SEQ ID NO 62; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This

CC sequence represents one of the isolated P. luminescens proteins

XX	SQ	Sequence 2410 AA;	
		Query Match 4.3%; Score 372; DB 6; Length 2410;	
		Best Local Similarity 20.0%; Pred. No. 1.8e-10;	
		Matches 307; Conservative 246; Mismatches 541; Indels 439; Gaps 75;	
QY		225 IEESLNKITTANNNDIRNLEKFADEDLVRLYNCELVERWNLAASDILRISMUK-----278	
DB		288 IEKTLTHLQTRGVVH-----NPALLSQALRDSSELKGTLPAM 325	
QY		279 -----EDGGVLDVILPGIOPDLF---KSINKPDSITNTSW-----EMIKLEAINMYKE 325	
DB		326 SGSEPRQVSEANTHQF-AADPVLFAKSENTINTETLRSGSPFAEGNAQLVKIESGLYELE 384	
QY		326 Y-----IPGVTSKNFMDLSEVORSFESALSSKSKSEIFLPLDDIKVSP 370	
DB		385 YTRSANDDDVAAGANSIPAY-----FLD-----YDSA-----DQANASP 418	
QY		371 LEVKIAFANNVINOALISUKSDYDLVINOIKRYKILND-NLNPISINEGTFNTWK 429	
DB		419 AYVDI--PKAIAESFLTCTLGGSVVVTSLDANTYRVYHGRVNSSL-----465	
QY		430 IFSDKLASISNEDNMFMKITYNLVKVPAPDVRSTINLSGPGVYTGAYQDILLMFKNST 489	
DB		466 LYDNVNVAVDDKYR--VAATDSLAVAYMYND-----EMQLVPQROYQO 511	
QY		490 NIHLEPELRFNFFP---KTKISOLTEQETLSWFSNQ-----ARAKSQFEEYKGYFE 540	
DB		512 DGQRLRLKRGDEEPLSIQIADSOVTEBNQAQATYREQVHQQLKKVATQFGVSVEGIAD 571	
QY		541 GALGE---DDNLDFAQNTVLDKDYVSKILSMK-----TRNKEY-----IHYLVQIQ 595	
DB		572 GVTTEGFSFDPHPAIAAWTKLHAE-VDDKINADIKQLVDRSKLYEEQRNTSOHNLIDQO 630	
QY		586 GDKIS-----YEASCN-----LFSKDPYSSILYQKNI-----EGS 615	
DB		631 IKQINITLEYKQAQNTVLEAGFVEQSWLMOQIKSGDKGAAVVR-DNTPIQNSGLEHTS 690	
QY		616 ETAYVYVVADEAIKIDKIRIPYQISNKRNIKLTFIGHKSSEPTDTFANLVDLSLSBI 675	
DB		691 SIGERYVIADA-----YQ-----HRVHGTAFTDGGQNFEDSIPGM-YDRNSA-L 733	
QY		676 ETILNLAKADISPK-----YIEINLLGCMNMFYSI-----SAEETYPGK 714	
DB		734 EMKILFLKGLTPKQQSALSAHITETTSQAEYIDKVLKWTAVYSEDTHRAGSSFDRLPQD 793	
QY		715 LLLKIKDRVS--ELMPSISQDSITVSANQYEVINEEGKREIL-----DHSGKWINKESI 768	
DB		794 FYLSLVGDKSGGRCYPLVRAMAVAL-ANSGEAGINNLVEXKFLAAADPHAGSSTLLKNSL 852	
QY		769 IK-----DISKEYIS-----FNPKENKIYKS-----KY 793	
DB		853 IKLHSDNLEAVQASTTRGQVTLSEWLLKATKTSTFALTNTFNHSMVMGSMVGLGTRY 912	
QY		794 -----LHELSTLLOEIRNA-----NSSDIDLEK--KVM 820	
DB		913 YFDPNVGIGFAFNNTSFLKAMKHLVGRKLAHYGLGDKLTFINLVEIDASKAKVP 972	
QY		821 LTECEINVAGNIDR-----QIVEGR--IEEAKN-----LTSDSINYIKNEFKLES-----864	
DB		973 VGN-DLVA-DLTPEDLTTVIEGRQVQEQAVNTQVRIAKDT--RLHTALATLDAEQWGS 1028	
QY		865 -ISDSLYDLKHQNGLDSPHSFIDSKTEN-----GPRIRPKNE-----TGNSTF 910	
DB		1029 RFAEASTRIAQENNLQDQW--MPVIGNTEDQKGHYRIQFINRDQPEETRWSSGDATF 1085	
QY		911 IETEKEIFSEVAT---HISKEISNIKDTIFDNVNGKLKVNLDAAHEVNTINSAPFIQS 967	
DB		1086 IEFRRFVDEHVLITGQHTFE-----HGQIQPGKGIGEAAPVDGLNAGFAIOS 1133	
QY		968 LIEYNTTKE-----SLSNLSVAMKVQVVAQLFSTGLNTITDASKVVELVSTAL----1015	

DB		1134 LIOWFTDKIRNDAARGIASPDLATALKVHSYLNVLQWMAHGGVQVTKVTELVTALRGEV 1193	
QY		1016 -----DETIDLPTLSSEGLPIIATIIDGVSLGAAIKSELSETNDPLLOEAEKIGIM 1067	
DB		1194 VAATSUKDFTSLIGHTVNEGA-----GVLPFGAMVGLDAYE--LAHAENDVQKAVF 1243	
QY		1068 AVNLTAAASTAIVTSALGIASGF-----SIL---LVPLAGISAGIFSLVNNELILODKA 1117	
DB		1244 GTQLAFDSASFTVGTAGVGAGLVGASTAGVILGGVAVGFTALACAFGVAEDA 1303	
QY		1118 TKVIDYFKHISLAETGAFTLLDD-KLIIMPQDDVLSEIDFNNNNSITLKGKCEIWRASGGS 1176	
DB		1304 KAVGRYPTIDKAYKNGRYDYNEKKVLVPLAGAVIKTDLRKHQIDFDSQYIYRTHSGS 1363	
QY		1177 GHTLTDDIDHFP-----SSPSITVRKPKWLSIYDVNLINIKKEKIDPS-KDLMVLN 1224	
DB		1364 --TSGGRINYPFWTGDFFKXVDRSQALEVRS-CIGYKDV-----SHKLDHSDSNVLILPG 1416	
QY		1225 APNRVFGVEMGWTGFRSLDNDGTLDRIRDHYEGQFYWRYPAFIADALITLKPRYED 1284	
DB		1417 TPKSFISYEYIQLSATRHDAGFDVIRLEE--DKRFDYDFYIFPGEVITIRRHQEWMA 1474	
QY		1285 TNVRINLDGNTRSFIVPVITTEQIRKNLSYFYGGSGYSLSLS-----PYNMIDNLV 1339	
DB		1475 TPIEWLDQRNRQLIVPELPKE-LHGYLHYEIKGAGGEYLLIGLNASVSKLTSVDPHNLV 1533	
QY		1340 EN-DTWVIDVDNVVKVNTITESDIOQKEL-IEHLSKLNIEDNKIILNNHITINFYGDIN 1396	
DB		1534 GSAPSRWIIDSSQLAS---DSISVSKNRLVVGWVAVVELSDRKEQVL---VWACKGEVR 1586	
QY		1397 ESNRFISLFTFILDINIIEIDLVSYSYKILLSGNCMKLIENSDDIOQKIDHIGFNGEH 1456	
DB		1587 EVD-FSNLTTQVVSSE-----DASKMQVFGYSIE--OHLUSDIAK-----AH 1623	
QY		1457 QKYPYPSYIDNETKYN-----GFIDYSKKEGLEFT 1485	
DB		1624 QLHGQYAVVEN-YRYNERNVGRAFYDVAKDRLMFT 1657	
		RESULT 11	
		AAG82935	
ID		AAG82935 standard; protein; 5024 AA.	
XX		AAG82935;	
AC		AC	
DT		03-SEP-2001 (first entry)	
DE		S. epidermidis open reading frame protein sequence SEQ ID NO:2964.	
KW		Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination; endocarditis.	
OS		Staphylococcus epidermidis.	
PN		WO200134809-A2.	
PD		17-MAY-2001.	
XX		09-NOV-2000; 2000WO-US030782.	
PR		09-NOV-1999; 99US-0164258P.	
PA		(GLAX) GLAXO GROUP LTD.	
XX		Kimmerly WJ;	
PI		WPI; 2001-316495/33.	
DR		N-PSDB; AAH53785.	
XX		Nucleic acids encoding polypeptides from Staphylococcus epidermidis, useful for vaccinating against infections, e.g. endocarditis.	

PS Claim 18; Page 779-781; 2188pp; English.

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAH81454 to AAH83120, from *Staphylococcus epidermidis*. (I)
 CC and (II) can have antibacterial activity and therefore can be used in
 CC vaccination. The nucleic acids (I) may be used to produce the S.
 CC epidermidis polypeptides (II) via the production of vectors containing
 CC them which are used to produce host cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55030 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55031 to AAH55098
 CC represent oligonucleotide sequences and primers which are used in the
 CC exemplification of the present invention. N.B. The present invention
 CC specifically claims all the polynucleotide sequences given in the
 CC sequence listing of the present specification, however the sequence
 CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
 CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
 CC for SEQ ID NO:4455 to 4464

XX SQ Sequence 5024 AA;

Query Match 4.1%; Score 355.5; DB 4; Length 5024;
 Best Local Similarity 19.6%; Pred. No. 3.4e-09;
 Matches 369; Conservative 310; Mismatches 626; Indels 579; Gaps 101;

QY 20 QDEYVAILNALEEYHNSESVEKYLKLDINN-----LTDN-----YLN-- 61
 DB 2611 QMD-----TILNHI--FSAPTRSGVGEKIASAKQLNNTMKALRDSIADNEILQSSKYFNE 2665

QY 62 TVYKSGRNKALKKFEYLTMEVLEKNNSLTPV-----EKNLHFIMIGG 105
 DB 2666 SEQQAQYNAQVNAKNI-----NDQPTFVMADEIQSVLNEVKTKDNLH-----GD 2713

QY 106 Q--IND-----TAINYNQWKNVDSYTVKVFYDYN-----AFLINTLKKT 144
 DB 2714 QKLANDKTDQAATNALNAYLNOAQRGNLETKVQ-----NSNSRPEVQKVQVQLANQLNDAMKK 2770

QY 145 IVESATNNTLSEFRENLDNPEFDYKFKRMEIYIDYKQHFIDYKQSIIE--NPEF-- 200
 DB 2771 LDALTGN--DAIKQTSYINEDTSQ-----QVNFDE---YTRGRKNVAEOTNNPNSP 2819

QY 201 -----IIDNIIKTYLSNYSKDEALNKYIBESLNKITANNNDIRNLEKPADDLVRLY 255
 DB 2820 TWINTIADKI-----TEAKNDLHVQKL--EQAQQQSININQMTGLNQAKQEQ-- 2867

QY 256 NOELVERWNLAAASDILRISMLKEDGGVYLDVILPG:QPDLFKSINKPDSITNTSWEMI 315
 DB 2868 NGEIQOTQTRSEVHVQVINKAALNDS-----WNTLRQSITDEHEVKOTSNVIN 2915

QY 316 KL-----EAIMKYEIPYTSKNFMDLDEEVQSFESALSKSKSIFLPLDDIK 367
 DB 2916 ETVGQGTATNNAVDVKQIQTINQTSNPTWNL--EVERATSNVTKTSK-DALHGERELNDNK 2972

QY 368 VSPLEVKLAFANVS--INQALISLSDSYCSDL-----VINQIKREYKTLN----- 411
 DB 2973 NSK-----TFAYNHLNLDNQA--QKEALTHEIQATIVSQVNNIYNKAKALNNDMKKLK 3024

QY 412 -----DNLNPSINEGTDFNTMTKIFSDKLASISNEDNMWFI-----KITNLYKVG 457
 DB 3025 DIVAQODNVQSNVYNEDSTQNNYND--TINHAQSIIDQVANTPMTSHDEIENAIN-- 3079

QY 458 PAPDVRSTIN-LSGPGVYTGAYQDILLMPKDNSTNTHLEPELRNPEFP-KTKISQLTE-- 513
 DB 3080 ---NIKHAINDLGE-----EKLQAKENA--NLLINSINDLNAPORDAINRLNVEA 3126

QY 514 -----QBITSLWFSNQA-----RAKQF-----EYKKGYPFALGEDDNL 549
 DB 3127 QTRKVAEQLOQAALNDAMKHLRNSIQNSSVROESKYINASDAKQCYNHAREVENI 3186

QY 550 DPAQNTVLDKDYVSKKILSSMKTRKEYIHYIVLQGGDKISYEASC----- 595
 DB 3187 INEQPTLDKEII--KQLTDAVNQANN--LNGVELLDADQKONAHQSIPTLMLHNOAQONAL 3244

QY 596 -----NLSKDPYSILYQ-----KNIGSETAYYYYVADAB-IKEIDKIRPIYQIS 641
 DB 3245 NEKINNNAVTRAKVAALIGQAKILDHAMENLBES-----INDKEQVQSSNY-----IN 3292

QY 642 NKENIKLTF--IGHKSEFFNTDTANLDVDSLSSEIETILNLAKADISPK---YIEINL 695
 DB 3293 EDPDVQETVNNAVDHVTEILNQTVMNPTSIDIEHAINEV-NOAKQRLRGKQKLYQTIDL 3351

QY 696 LGCNMFYSISABETYPGKLLKIKDRVSELMPSISQDSITVSANQYEVRIEENKGR--- 752
 DB 3352 -----AKELSKLDDLSLQSSSSISNQIYIAKTRTEVAQAIEKAKSLNH 3395

QY 753 -----EILDHSGKWINE-----ESIIMDISKEYISFNPKENKIIVSKK 792
 DB 3396 AMKALNKIYKNAKVDL--SSREINDEQEKAYQAOAHNVDSIIHROTNPEMDPTVINS- 3453

QY 793 YLHELSTLLOETRN-----ANSSDDIDLEKXWMLTECEIN-----VASNI 832
 DB 3454 ITHELETAQNHLHGDKLAHAKQDAANVINGLIHLNVAQREVMINTNTNATTREKVAKNL 3513

QY 833 DR-QIVEGRIEBAKNLSDSINYIKNEPKLIESDSLYDLKHONGLDDSHFISPED:SK 891
 DB 3514 DNAQALDKAMETLQOVVAHK--NNILNDSKYLN--EDSKYQQQYDRVIADAEOLLNQTNP 3570

QY 892 TENGFRIRPINKETGNSPIETEKEIFSEYATHISKEISNIDKTIFFDNVNGKLVKKNLD 951
 DB 3571 TLEPKYDVIVKDNV-----LANEKILFG--AEKLSYDKSNANDEI----- 3608

QY 952 AAHEVNTLNSAFPIOSLIEYNTKESLNSLVAMKVQVYQLFST---GLNITDASKV 1008
 DB 3609 --KHMNYLANA-----QKSIKDMISHAALTEVVKQLLQQAKTLDKAMKLEDKTVV 3659

QY 1009 ELVSPAL-----DETIDLLPTLSEGLPIIATII-DGVSLG--AAIKELSETNDP 1054
 DB 3660 -ITDTPLPNYTBASEDKKEKVD--QTVSHAQALIDKINGSNVSLDQVQALFQASEN 3716

QY 1055 L--LRQETAEKI-----GIMAVNLTAASTA--IVTSALGIASGFSILLVPLAGISAG 1102
 DB 3717 LDGQQRVE-EAKVHANQTIQDLTHLNS-LQQQTAKESVKNATKLE----- 3760

QY 1103 IFSLVNNELIQDKATKVIDYFKHSLAETGAFTLLDDKIIIMPQDDLVLSFIDFNNSI 1162
 DB 3761 IATASNALALANKWKLQFQFINHADSIEANSYRQADDKIIAYDDALEHQDQIKSNA 3820

QY 1163 TLGKEIWEAEGSGHTLTDIDHF-FSSP-SITRKPKWLSIYDVNLNK-----EKIDPS 1216
 DB 3821 TQNEAKALQALQALINAETSLNGPERLNHAPRALEYIK---SLEKINNAQSALEDKVTOS 3877

QY 1217 KDLMLVLPNAPNRVFGYEMGWTGFRSLDNDGTKLLDRIDHYEGQFYWYFAFIADALIT 1276
 DB 3878 HDLLELEH-----LVNEGINDIMGE-----LANAIVN 3906

QY 1277 KLKPRYEDTNVRINDGNTRSPFIVPITTEQIRK-NLSYFSGSGSGSYSLSPYNNMID 1335
 DB 3907 ---NVAPTKASIN-----YINADNRKONFTQAINNARDALN---KTQGNULD 3948

QY 1336 LNLVENDTVIDVD-----NVVKNTIDESIQKGLIENI--LSKMI-EDNKIILKN 1386
 DB 3949 FNAI--DTFKDDIFKTKDALNGIERLTAAKSKAEK--LIDSLEKFNKAQFTHANDEIMNT 4004

QY 1387 HTNFGVDNENRFLSIFSLFSDILEINIIEIDLVSKYSKILLSGNCKMLIENSSDIQOK 1446
 DB 4005 NS-----IAQLSRIVNQAFD-LNDAMKSLRDLANQAFVQOASN---YINSDDELKQO 4054

QY 1447 IDHIGFNGEHQKYPYYSIDNETKNGFYDYKGEGLFTAEFESNESIIRNIYMPDSNNLF 1506
 DB 4055 FDBA-----LSNARKV-----LAKENG----- 4071

QY 1507 IYSSKDLKDIRIINKGDVKLLIGNVFKDDMKVSLFTIEDTNTIKLNGVYLDENGVA--- 1563

Db 4072 -----KNLDEIQI-----EGLKQVIEDTKD-ALNGI:QLRSKAKAKAI 4107
QY 1564 ---QILKPMNAKSALENTSN-----SLMFELESINIKIYNNLDNIEFLDNTFI 1612
Db 4108 QYQSLSYINDAQRIHAESIHNSDDLSSLANLSKA-----SOLDNAMKDLRDT--L 4158
QY 1613 ISGNSISQPELICDKDKNTQPYF 1636
Db 4159 ESNSTSVPNVSNVINADKNLQIEF 4182

RESULT 12

ABU42656
ID ABU42656 standard; protein; 6641 AA.

AC ABU42656;

19-JUN-2003 (first entry)

Protein encoded by Prokaryotic essential gene #28183.

Antisense; prokaryotic essential gene; cell proliferation; drug design.

Staphylococcus epidermidis.

WO200277183-A2.

03-OCT-2002.

21-MAR-2002; 2002WO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

WPI; 2003-029926/02.

N-PSDB; ACA46526.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 70580; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 5213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies on a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 6641 AA;

Query Match 4.0%; Score 347; DB 6; Length 6641;
Best Local Similarity 19.4%; Pred. No. 1.4e-08;
Matches 354; Conservative 318; Mismatches 662; Indels 494; Gaps 97;
QY 20 QEDYVAILNALAEYHNHNSSESVVEKYLKLDINN-----LTDN-----YLN-- 61
Db 3989 QKD---TILNHI--PSAPTRSQVGEKIASAKQLNNTWKALRDSIADNNEILQSSKYFED 4043
QY 62 TYKSGRNKALKKKEKYLTMVELEKNSLTPV-----NQDPTVWANDIEIQSVLNEVKQTKDNLH---GD 4091
Db 4044 SEQQAYNQAYNAKKNII-----NDQPTVWANDIEIQSVLNEVKQTKDNLH---GD 4091
QY 106 Q--IND-----TAINYNQWQVNSDYTVKVFYDSN-----AFLNLTAKT 144
Db 4092 QKLANDKTDQAATNALNLYNQAGRNLETQV---NSNSRPEVQVKVOLANQNDAMKK 4148
QY 145 IVESATNNTLESFRENLDPEFDYNNKFKYRMEIYDKQHFIDYKQIBE--NPEF-- 200
Db 4149 LDDALTGN--DAIKQTSNVINEDTSQ-----QVNFDE--YDRGKNIVAEQTPNNMSP 4197
QY 201 -----IIDNIKTLYLSNYSKDLALNKY-----IESLNKIYANNNDIRNLEKFADELDV 252
Db 4198 TNNTIADKI-----TEAKNDLHGQVKLQQAQCSINTI-----NOMTGLNQAKQEL- 4245
QY 253 RLYNQELVERWNLAASDILRISMLKEDGGYVLDVILPGIQPLFKSINKPDSITNTSW 312
Db 4246 ---NOEIQOTQTRSEVHQVINKQAALNDS-----MNTLRQSIITEHEVKQTSN 4290
QY 313 EMIKL-----EAIMKYKEYIPGYTSKNFTMDLDEEYQVSFESALSCKSKSEIFLPLD 364
Db 4291 YINETVGNQATYNNADVQKVIQNTSPTWNPPL--EVERATSNVKISK-DALHGERELN 4347
QY 365 DIKVSPLVYKIAFANNSV--INQALISLSDSYCSDL-----VTNQKNRYKILN---- 411
Db 4348 DNKSK-----TFAVNHLNLAQA---QKEALTHEIEQATVSVQVNNIYNKAKALNNDMK 4399
QY 412 -----DNLNPSINEGTDFNTMKIFSDKLASISNEDNMFM-----KITNYL 454
Db 4400 KKLQDIVAQDNVRSQNNYINEDSTPQNMVND---TINHAQSIIDQVANTPTMSHDIEENAI 4456
QY 455 KYGFPADYRSTIN--LSGFGVYTGAYQDILLMPKDNSTNIHLEPELRNFEFP-KTKISOLT 512
Db 4457 N-----NIKHANLADGE-----HKLQQAQKENA---NLLNSLNDLNAPODAINRLV 4501
QY 513 E-----QBITSLWSFNQA-----RAKQSF---EYKKGVFEGALGED 546
Db 4502 NEAQTRFKVAEQLQQAALNDAMKHLRNSIQNSSVRSQSKYINADAKKEQYNHAREV 4561
QY 547 DNLDFQANTVLDKDYVSKKILSSMKTRNKKEYIHYIVQLQGDKISYEASCN--LFSKDPYS 604
Db 4562 ENLINEQHPTLQKEII-KQLTDGVNAQAND-LNGVELLDADKQNAHQSIPTLMHLNQAOQ 4619
QY 605 SILYQK-NIEGETAYYYVADAETKEIDKTRIPQISKNRNIKLFTFHGSKSEFTDTF 663
Db 4620 NALNEKINNNAVTRTEVAAILIQAKLLDHAMENLEESIKDKEQVK-----QSSNYINEDSD 4674
QY 664 ANLVDVDSLSEIETILN-----LAKADISPKVIRIN-----LLGCNMFYSISAEETYP 712
Db 4675 VQETYNDAVDHVTETILNQTNTPTSLIEDIEHAINVQNAQKQLRGKQKLYQTIDLAD---- 4731

Qy	713	GKULLKIKORVSELMPSISODSITVSGANQVEVRINBEGKR-----EILD	756
Db	4732	-KELSKLDLITSOQSSISINQIYTAKTRETEVAQAI EKAKSLNHAMKALNKVYNADKVID	4790
Qy	757	HSCKWINKE-----ESIIKD:SSKEYISFNPKENKIIVKSYLHLSHELSTLLQERNN--	807
Db	4791	-SSRFINEOPEKKAQOALNHVDSIIHROTNPEMDPTVINS-ITHELETAQNNLHGDK	4848
Qy	808	-----ANSSDIDLEKKVMLTECEIN-----VASNIDR-QIVEGRIEEAKNJT	848
Db	4849	LAHAQQAANVINGLIHLNVAQEVMTNTNATTREKVAKLNDAQALDKAMETLQQVV	4908
Qy	849	SDSINYKNEFKLIESDSLYDLKHQNGLDHSHFISFEDISKTENGFRIRPINKETGNS	908
Db	4909	AHK--NNILNDSKYL--EDSKYQQQYDRVIADAEQLLNQTNTPLEPQYKVDIVKQNV--	4962
Qy	909	IFETEKEIPESEYATHISKEISIMIKOTIPDNVNGKLVKKWLDAAEVNTLNSAFIQL	968
Db	4963	--LANEKILFG--AEKLSYDKSNANDEI-----KHMVNLNA-----	4995
Qy	969	IEYNTTKESLNSIVAMKVQ--VYAAQFSTGLNTITDASKVVVELVSTAL-----	1015
Db	4996	-QKQSKIDMISHAALRETVKQLLQQAQKILDEAMKSLEDKTVQV-ITDITPLNPYTEASEDK	5053
Qy	1016	DEITDILLTLPSEGLPIIATII-DGVSLG---AAKELSEINDPL---LRQEIEAKI----	1064
Db	5054	KEKVD--QTVSHAQAIIDKINGSNVSLDQVRQALEQLTOASENLGDGQORVE-EAKVHANO	5110
Qy	1065	---GIMAVNLTAASTA--IVTSALGIASGFSILLVPLAGISAGIPSLVNNELILQDKATK	1119
Db	5111	TIDOLTHLNSLQOQTAKESVKNATKLEE-----IAVSNNAQALNKVMGK	5155
Qy	1120	VIDYFKHISLAETEGATLLDDKLIIMPQDLVISEIDFNNSITLKGCEIWRABGSGHT	1179
Db	5156	LEQFINHADSVENSNDYRQADDDKIIAYDEALEHGGQIOKTNATQNETK-----	5204
Qy	1180	LTDDIDHFFSPSITYRKPLMSIYDVNL-1KKEKIDFSKDLMLPNAPNEVFEGVMGWP	1238
Db	5205	-----QALQQIYVATSELSINGERLUNHARPALEYIKSLEKINNAQKSALEDKVTQSH	5256
Qy	1239	GFRSLD--NDGTKLLDRIDHVBEGQFYWRYPAFIADALITKLKPRYEDTNVRLDNGT	1295
Db	5257	DLLEHIVNEGTNLNDIMGSE-----LANAIVN---NYAPT-KASIN-----	5294
Qy	1296	RSFIVPVITTEQIRK-NLSVSYFGSGSYSLISPYNMNIDMLNVENDTVIDVD-----	1349
Db	5295	-----YINADNLKDRFTQAINNARDALN--KTQGNLDFNAI--DTEFKDDIFKTKDA	5343
Qy	1350	NVKNKITESDEIKQGELIENI--LQKLNLI-ENKILNHTINFYGDINEGNRPISLT	1405
Db	5344	LNGHERITAAKSAEK--LIDSLKFINKAQFTHANDEIINTNS-----IAQISRVNQOA	5395
Qy	1406	PSILEDINIIEIDLVSQSKYILLSGCMKLIENSDDIOQKIDHIGFNGSHQKYPYSYI	1465
Db	5396	FD-LNDAMKSLRDELNQAAPVQASSN---YINSEDEKLQOFHALSNA--RKVL-----	5444
Qy	1466	DNEPKYNG-FIDYKSKKGLFTABESNESIERNIYMPUSNLIIFYSSKD-LKDRIINKGD	1523
Db	5445	---AKENGKNLDEKQIQGL-----KQVIEDTKDALINGQRUSKAK	5481
Qy	1524	VKLLIGNYFKDMKVSLSFTTIEDTNTIKLNGVYLDENGVAQILKFMNNAKSALNTSLSM	1583
Db	5482	AKAI--QYVQ-----SLSYINDAQRHAEANNIH-NSDDLSSLANTLSKASDLONAMKDLR	5533
Qy	1584	NFLSISINI---KNIFYNMLDPNIEFILD	1608
Db	5534	DTIISNSTSVPSNVYINADKNLQIIEFD	5561

RESULT 13

ABP38314
ID ABP38314 standard; protein; 10182 AA.

AC	ABP30314;
XX	
XX	24-JUL-2002 (first entry)
XX	
DE	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3159.
XX	
KW	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX	antibacterial; gene therapy.
XX	
OS	Staphylococcus epidermidis.
XX	
PN	US6380370-B1.
XX	
PD	30-APR-2002.
XX	
PF	13-AUG-1998; 98US-00134001.
XX	
PR	14-AUG-1997; 97US-0055779P.
XX	
PR	08-NOV-1997; 97US-0064964P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Doucette-Stamm LA, Bush D;
XX	
DR	WPI; 2002-381255/41.
XX	
DR	N-PSDE; ABN90859.
XX	
PT	Novel isolated nucleic acid encoding a Staphylococcus epidermis
XX	polypeptide, useful for diagnosing and treating bacterial infections.
XX	
PS	Disclosure; SEQ ID NO 3159; 267pp; English.
XX	
CC	ABN905338 to ABN93374 represent Staphylococcus epidermidis open reading
CC	frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC	given in ABP35124 to ABP7960. The S. epidermidis sequences have
CC	antibacterial activity and can be used in gene therapy. The sequences can
CC	also be used in the diagnosis and treatment of bacterial infections,
CC	particularly S. epidermidis infections. The sequences can be used to
CC	screen for compounds able to interfere with the S. epidermidis life cycle
CC	or inhibit S. epidermidis infection. N.B. The sequence data for this
CC	patent did not form part of the printed specification, but was obtained
CC	in electronic format directly from the USPTO web site
XX	
SO	Sequence 10182 AA;

QY 253 RLYNQELVERNNLAASDIIRISMLKEDGGVYLDVILPQIOPDLFKSINKPKDITNTSW 312
|||:::||||:::||||:::||||:::||||:::||||:::
Db 7814 --NQEIQQTOTRSEVHQVINKAQLANDS-----MNTLRQSITDEHEVKQTSTN 7858

CC of infection with P. falciparum. Furthermore, (1) (especially when they
 CC are rifins or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic life cycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAA8144 to AAA81852
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification
 XX
 SQ Sequence 2485 AA;

Query Match 4.0%; Score 343; DB 3; Length 2485;
 Best Local Similarity 19.7%; Pred. No. 6.1e-09;
 Matches 356; Conservative 269; Mismatches 643; Indels 542; Gaps 91;
 QY 35 HNSSESVVYKYLKUDINNLTDN-----YLVTKYKGRKALKKF---KEYITMEVLE 85
 Db 157 HTNSNI---YINCNDNNDIRNSNVQHYNDKIKENKQNKYVLINDYNNKYI- 212
 QY 86 LKNSLTPEVKLHFIWIGQINDTAINVQ-----WKVNSDYTVKVFYDSNAFLN 139
 Db 213 LSKNKTCKINQKGL-----IKKKVANSRRNHILYKCRNLKLYNGVFSDD- 260
 QY 140 TLKKTIVGATN--NTLESFRENAN--DPEPDYKFKYKRMELIYDKQHFIDYKSOI 194
 Db 261 -----IKSEVNVCSLTVLHKNYINMDNYLDDN-----IHTNNSNIYD-----I 301
 QY 195 EENPEPIIDNIKT-----LSNEYSKDLALN--KYIEESLNKLTAN----- 235
 Db 302 NYTENVINSTCRYPIGNNTVLSXDEVTKSSSKNSLSYFDDIIN-VKNNDIPILHDK 360
 QY 236 NGNDIRNLEKF-----ADELVRLYNQELVER-----WNLAAADIIRISMLKEDGGVY 284
 Db 361 NINISNKSCHKEKEEKEEYKMYNSNLVEEKQKKMIWN-----SLNVLPIDLLKNG--- 414
 QY 285 LDVDILPGQPDLFKSINKPDSITNTSMEMIKLBALMKYKEYIPGYTSKNFMDLDEEVR 344
 Db 415 -----HDEINKEICKKKKSPSQNDIKSKMLYNNKSY-----SKS----- 450
 QY 345 SPESALSSKSDSEIFPLDDIKVPLEVKIAFANNVINOALISLKDSYCSDLVI----- 400
 Db 451 --EKVLYTNKNSNTPIPIFFLN-----KYGDFKNSENIYDMYNNKQVYIHDKLIYTM 504
 QY 401 --NQIKORY-----KILNDNLNPSINEGTFDNTM--KIFSD-KLASISNEDNMWPMI 448
 Db 505 YSNKLAQKHVYTSNINLNNIGKVLNDLHLSNNMYCRLNSNPPYKSIISLNNNVFFY 564
 QY 449 KITNYLKVGFADVRSTINLSPGVYTGAYQDLLMFKD-NSTNIHLEPE---LNPFP 504
 Db 565 K-----KRKSNNSNNNNNNISSSSSSKQKHVIINKKISSYNIHYKPKDSFKNFLFF 620
 QY 505 KTKISOLTTQETISLWSFNQAKAQSFEBYKGYEGALGEDDNLDFQNTVLDKDYYSK 564
 Db 621 KEKI--LPSKDTCTV-----FNERQKDLFE----- 643
 QY 565 KILSSMKTNKKEYIHYIVQLQ--GDKISYEASNLFSKOPY-----SSILY---QKNIEG 614
 Db 644 -----KSENIHKVCSFNNTSDDISHSSVN--KKEPFFALKNSIRHIPPKNENIY 693
 QY 615 SETAYYYYVADAEIKDYRIPYQISNKNRTKLPIGHGKSEFNTDFANLDVDSLSSE 674
 Db 694 TSGKSNHVDKDEKTVLLKKK--KEINDKNTFSSCLINHITTYTLQNGVKN----- 744
 QY 675 IETILNLAKADISPKYIEINLGCNMFYSIS-----AEETYPKLLIKDQVSELNPS 729
 Db 745 -----LNMLGIRDSIYKIDKKNMLKECYNGNNDNNKKNKKKKKKKLS 786

QY 730 ISQDSITVSANOYEVRAINEEGKREILDHS-GKWINKESIIKDISKE-----YISFNP 782
 Db 787 FSCJLINDNITPVESDEKKNNSNIIKSMDFINVVKRKSNNLYNNLSNRDSTVDMENKYS 846
 QY 783 KENKIIVKYLHELSTLLOETIRNANSSDIDLEKKWMLTECEI--NVASNIDRQIVEGR 840
 Db 847 EEVNIQRTNKIVLSN--KRIRNYKLYS-MDEIFKVLKEKKYIDNISNNMER----- 897
 QY 841 IEBAKNLTSDSINYIKNEF--KLIESISDLYDLKHQNGLDSDSHFISFP-DISKTENGPR 897
 Db 898 -----VYI-KNEMINEKISKMDILLPCCKNKLNSMPCVPIIENNISREN----- 942
 QY 898 IREINKETGNSPIETEK-----EIPSEYATHISKEISNIKOTIFDNVNGK-----LVKK 947
 Db 943 -----EKNSSVILNKKKNENMENCVRHLCHWGK--YNNQDNIVDQGNIKNEEBEITKH 994
 QY 948 VNLDAAEHVNTLSA-----FFIQSLIEYNTTRESLSNLSVAMKVQVYAOQLFSTGLT 1000
 Db 995 DEYISREKKNYKSKIRNFDDYKYEQVLSYHTLDEDKKK----- 1034
 QY 1001 ITDASKVVELVSTALDETIDLLPTLSEGLPIITATII---DGVSLGAATKELSETNDPLL 1056
 Db 1035 -NDMNLIIDMNEAIIETYN-----GVNNIILDRKNNSRKDKMEKEMEKEMKKM 1084
 QY 1057 ROIEAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILOPK 1116
 Db 1085 EKEMEK--VMEKEMEKVMEKEVEKEL-----KNEMNNNNNNNNNNNNNNNNNN 1129
 QY 1117 ATK-----VIDYFKHISLAETEGAFITLDDKIIMPQDDLVLSIEDFN--NNSITLGCBEIWR 1171
 Db 1130 IYKXNEIYVDNDKELEIVNEE-----KKLIVFPN--YESDVHKOMMSINNNCK--- 1177
 QY 1172 AEGSGHTJTDIDHDFSSPSITYRKPWLSIVDVNLKKEKIDFSKDLMLVLPNAPNRVVG 1231
 Db 1178 -----DDVN-----NILKEYVDNS-----CLAQKEENI-- 1200
 QY 1232 YEMGWTFPSRLDNDCTKLLDRIRDHYEQGVYKFAF-----IADALIT 1276
 Db 1201 -----FRPLENKK--DKV-----WKRFNKNKNTIHIHEEMKRIYOTINK 1241
 QY 1277 KLKP-----RYED--TNVRINLDNGTRSFIVPVTITEQIRKNLSYSFYSGSGYSLSLS 1328
 Db 1242 NVFPIYNFREYENFLNHLTYNFPKNDLPKLSYKVSMMNIR-NLYTANKHINNIDYMNK 1300
 QY 1329 PYNNMI-----DLNLVENDTWI-----DVDNVVKNITIE----- 1358
 Db 1301 LYQNQIVTLKYQVANIDNDHHICKGGGLDYINNNISKECKNRKDKTYLKNIFHYKKKD 1360
 QY 1359 -----SDEIOKGELIENILSKLNIENDKILANH-TINFYD-----INESNRFISL 1404
 Db 1361 ARFFINDEIGSDNYMDYDIKKYSNDENNYKLANEKOMNISNDEDMIPTLNSHGNNFPSC 1420
 QY 1405 TFSILEDINIIEIDLVSYSKILLSGCMKLIENSSDIQQKIDHIGFNGEHQKYPYGY 1464
 Db 1421 QPNLLEKKSITYDNLNLYDSNMDDFTEBKYNFVNNDL-----FNTKWK---FNP 1469
 QY 1465 IDNETKYG-FIDYSKKEGLTAEFNSNESIIRIYMPDSNNLIYFYSKOLKOIRIINKGD 1523
 Db 1470 SKGKNLENKFFNVNSDGVFSF-FKNMNLFREL-NKSNNLSKLESVKNSNNNSNKG 1527
 QY 1524 VKLLIGNY-----FKDDMKVSL-----SFTIEDNTI--KNG---VYLDENG 1562
 Db 1528 DN--IGNMNNNTNVTIASDEHISTKGDHDSFSDRDNDCILLKIEGRSKKYS- 1581
 QY 1563 AQILKFNNAKSAINTSNLSMNFLESINIKNIPYNNLDNIEFILDTNFIISGS-NSIGQ 1621
 Db 1582 ---ITLYNEDKSLNENDNETINEYV-----CSNIDVN-----EWEDKVGTCNSVG 1627
 QY 1622 FELICDKDN 1631
 Db 1628 KETEKNNKN 1637

RESULT 15	Query Match	3.9%; Score 340.5; DB 6; Length 5005;
ABU48947	Best Local Similarity	19.7%; Pred. No. 2.1e-08;
ID ABU48947 standard; protein; 5005 AA.	Matches	405; Conservative 283; Mismatches 676; Indels 697; Gaps 105;
AC ABU48947;		
XX 19-JUN-2003 (first entry)		
DT 19-JUN-2003 (first entry)		
XX 19-JUN-2003 (first entry)		
DE Protein encoded by Prokaryotic essential gene #34474.		
XX 19-JUN-2003 (first entry)		
XX 19-JUN-2003 (first entry)		
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.		
XX 19-JUN-2003 (first entry)		
OS Ureaplasma urealyticum.		
XX 19-JUN-2003 (first entry)		
PN W0200277183-A2.		
XX 19-JUN-2003 (first entry)		
PD 03-OCT-2002.		
XX 19-JUN-2003 (first entry)		
PF 21-MAR-2002; 2002WO-US009107.		
XX 19-JUN-2003 (first entry)		
PR 21-MAR-2001; 2001US-00815242.		
PR 06-SEP-2001; 2001US-00948993.		
PR 23-OCT-2001; 2001US-0342923P.		
PR 08-FEB-2002; 2002US-00072851.		
PR 06-MAR-2002; 2002US-0362699P.		
XX 19-JUN-2003 (first entry)		
XX 19-JUN-2003 (first entry)		
PA (ELIT-) ELITRA PHARM INC.		
XX 19-JUN-2003 (first entry)		
PI Wang L, Zamudio C, Malone C, Haeelbeck R, Ohlsen KL, Zyskind JW;		
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;		
PI WPI; 2003-029926/02.		
DR N-PSDB; ACAS2817.		
XX 19-JUN-2003 (first entry)		
XX New antisense nucleic acids, useful for identifying proteins or screening		
PT for homologous nucleic acids required for cellular proliferation to		
PT isolate candidate molecules for rational drug discovery programs.		
XX 19-JUN-2003 (first entry)		
PS Claim 25; SEQ ID NO 76871; 1766pp; English.		
XX 19-JUN-2003 (first entry)		
XX The invention relates to an isolated nucleic acid comprising any one of		
CC the 6213 antisense sequences given in the specification where expression		
CC of the nucleic acid inhibits proliferation of a cell. Also included are:		
CC (1) a vector comprising a promoter operably linked to the nucleic acid		
CC encoding a polypeptide whose expression is inhibited by the antisense		
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated		
CC polypeptide or its fragment whose expression is inhibited by the		
CC antisense nucleic acid; (4) an antibody capable of specifically binding		
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular		
CC proliferation or the activity of a gene in an operon required for		
CC proliferation; (7) identifying a compound that influences the activity of		
CC the gene product or that has an activity against a biological pathway		
CC required for proliferation, or that inhibits cellular proliferation; (8)		
CC identifying a gene required for cellular proliferation or the biological		
CC pathway in which a proliferation-required gene or its gene product lies		
CC on a gene on which the test compound that inhibits proliferation of an		
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a		
CC product is overexpressed or underexpressed; (12) determining the extent		
CC to which each of the strains is present in a culture or collection of		
CC strains; or (13) identifying the target of a compound that inhibits the		
CC proliferation of an organism. The antisense nucleic acids are useful for		
CC identifying proteins or screening for homologous nucleic acids required		
CC for cellular proliferation to isolate candidate molecules for rational		
CC drug discovery programs, or for screening homologous nucleic acids		
CC required for proliferation in cells other than <i>S. aureus</i> , <i>S. typhimurium</i> ,		
CC <i>K. pneumoniae</i> or <i>P. aeruginosa</i> . The present sequence is encoded by one of		
CC the target prokaryotic essential genes. Note: The sequence data for this		
CC patent did not form part of the printed specification, but was obtained		
CC in electronic format directly from WIPO at		
CC ftp.wipo.int/pub/published_pat_sequences		
XX 19-JUN-2003 (first entry)		
SQ Sequence 5005 AA;		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:37:55 ; Search time 38.25 Seconds
(without alignments)
8413.819 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1020

Perfect score: 5189

Sequence: 1 MNLVNAQLQKVVYKFRQ.....ITDASKVELVSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriaph.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5189	100.0	2364	2 Q46342	Q46342 clostridium
2	4105.5	79.1	2367	2 Q9EXR0	Q9EXR0 clostridium
3	4096.5	78.9	2367	2 Q9F931	Q9F931 clostridium
4	4088.5	78.8	2367	2 Q46034	Q46034 clostridium
5	2156.5	41.6	554	2 Q93L39	Q93L39 clostridium
6	1916	36.9	698	2 Q68653	Q68653 clostridium
7	1898	36.6	698	2 Q9EXQ8	Q9EXQ8 clostridium
8	1892	36.5	697	2 Q86141	Q86141 clostridium
9	1474	28.4	553	2 Q9FCX5	Q9FCX5 clostridium
10	1435.5	27.7	2178	2 Q46149	Q46149 clostridium
11	422.5	8.1	3169	2 Q9ZGR4	Q9ZGR4 escherichia
12	419.5	8.1	3169	2 Q82916	Q82916 escherichia
13	391	7.5	3223	2 Q8L1L9	Q8L1L9 escherichia
14	389	7.5	3223	2 Q9RPH1	Q9RPH1 escherichia
15	386	7.4	3223	2 Q9RM48	Q9RM48 escherichia
16	386	7.4	3223	2 Q8VQR2	Q8VQR2 escherichia

17 382 7.4 3223 2 Q8VNR6
18 358.5 6.9 2771 5 Q26216
19 356 6.9 3346 16 Q822X0
20 352 6.8 3317 16 Q8EWP8
21 350 6.7 3225 16 Q9PKM6
22 348.5 6.7 2752 5 Q9BJY0
23 345.5 6.7 1127 12 Q9YVT6
24 345 6.6 5767 5 Q81525
25 343 6.6 2965 5 Q8WR55
26 342 6.6 2770 5 Q7YUE9
27 341 6.6 2976 5 Q8WR54
28 338 6.5 2204 5 Q8IL70
29 337.5 6.5 2867 5 Q9N2M3
30 336.5 6.5 2747 5 Q9BJX9
31 335.5 6.5 2957 5 Q8WR56
32 328 6.3 3724 5 Q77320
33 321.5 6.2 1238 12 Q9EMP3
34 319.5 6.2 2166 16 Q51465
35 318.5 6.1 3394 5 Q77384
36 317.5 6.1 3096 5 Q7YWP2
37 317.5 6.1 3254 5 Q9BK45
38 317.5 6.1 3254 5 Q7YWE9
39 316.5 6.1 3081 5 Q7YWF1
40 316.5 6.1 3203 5 Q7YWE7
41 316.5 6.1 3203 5 Q7YWE6
42 316.5 6.1 3255 16 Q9PKM8
43 316 6.1 1958 5 Q8L1E1
44 316 6.1 3256 5 Q7YWE8
45 315.5 6.1 1387 5 Q9GZ76

ALIGNMENTS

RESULT 1

Q46342 PRELIMINARY; PRT; 2364 AA.
AC Q46342
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Cytotoxin L.
OS Clostridium sordellii.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1505;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6018;
RX MEDLINE=95369733; PubMed=7642137;
RA Green G.A., Schue V., Montell H.,
RT "Cloning and characterization of the cytotoxin L-encoding gene of
RT Clostridium sordellii: homology with Clostridium difficile cytotoxin
RT B.",
RL Gene 161:57-61(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=6018;
RX MEDLINE=96149194; PubMed=9544213;
RA Green G.A., Schue V., Girardot R., Montell H.,
RT "Characterisation of an enterotoxin-negative, cytotoxin-positive
RT strain of Clostridium sordellii.",
RL J. Med. Microbiol. 44:60-64(1996).
DR EMBL; X82638; CAA57959.1; -
DR PIR; I40884; I40884.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR GO; GO:0006413; P:translational initiation; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR007577; Gly_transf_sug.
DR InterPro; IPR001950; TIF_SUI1.
DR Pfam; PF01473; CW_binding_1; 18.
DR Pfam; PF04488; Gly_transf_sug; 1.
DR PROSITE; PS01118; SUI1_1; 1.

SQ SEQUENCE 2364 AA; 270576 MW; EAD8A4467A89BDBB CRC64;

Query Match 100.0%; Score 5189; DB 2; Length 2364;
 Best Local Similarity 100.0%; Pred. No. 3.4e-190;
 Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNLVNAQLOKQVYKFRIOEDYVAIINALEEYHNMSESSVVEKYLKLDINNLTDNYL 60
 DB 1 MNLVNAQLOKQVYKFRIOEDYVAIINALEEYHNMSESSVVEKYLKLDINNLTDNYL 60

QY 61 NTYKSGRNKALKKPKFYLTMVELEKNSLTPVEKNLHF*WIGQINDTAINYNQWKD 120
 DB 61 NTYKSGRNKALKKPKFYLTMVELEKNSLTPVEKNLHF*WIGQINDTAINYNQWKD 120

QY 121 VNSDYFVKVYDSNAPLINTLKTIVESATNTLTSPRENLDPEFDYKFRMRMIY 180
 DB 121 VNSDYFVKVYDSNAPLINTLKTIVESATNTLTSPRENLDPEFDYKFRMRMIY 180

QY 181 DKQKHFIDYKSOIENPEFIIDNTIKTYLSNEYSKDLEALNKYIEESLNKITTANNNDI 240
 DB 181 DKQKHFIDYKSOIENPEFIIDNTIKTYLSNEYSKDLEALNKYIEESLNKITTANNNDI 240

QY 241 RNLEKFAEDLVRNLQELVERWNLAAASDIIRISMLKEDGGVLDVDILPGIOPDLFKS 300
 DB 241 RNLEKFAEDLVRNLQELVERWNLAAASDIIRISMLKEDGGVLDVDILPGIOPDLFKS 300

QY 301 INKPDSITNTSWEMIKLEALIMKYKEYIPGYTSKNPDMLEDEEVQSFESALSSKSDKSEIF 360
 DB 301 INKPDSITNTSWEMIKLEALIMKYKEYIPGYTSKNPDMLEDEEVQSFESALSSKSDKSEIF 360

QY 361 LPDLDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINQIKRKYILNDNLNPSINE 420
 DB 361 LPDLDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINQIKRKYILNDNLNPSINE 420

QY 421 GTDFNTMTKIPSDKLASISNEDNMFMFKITNLYKVGFPADVRSTINLSGPGVVTGAYQD 480
 DB 421 GTDFNTMTKIPSDKLASISNEDNMFMFKITNLYKVGFPADVRSTINLSGPGVVTGAYQD 480

QY 481 LLMFKDNSTNIHLLEPELRNPFPTKISQLTEQBITSLWFSNQAQAKSPPEYKKGYPE 540
 DB 481 LLMFKDNSTNIHLLEPELRNPFPTKISQLTEQBITSLWFSNQAQAKSPPEYKKGYPE 540

QY 541 GALGEDNDLPAQNTVLDKDYVSKILSSMKTRNKEYIHYIVQOGDKISYEASCNLF 600
 DB 541 GALGEDNDLPAQNTVLDKDYVSKILSSMKTRNKEYIHYIVQOGDKISYEASCNLF 600

QY 601 DPYSSTLYOKNEGSETAYVYVADAETKEIDKYEIPYCIENKRNKILTFTHGKSEFNT 660
 DB 601 DPYSSTLYOKNEGSETAYVYVADAETKEIDKYEIPYCIENKRNKILTFTHGKSEFNT 660

QY 661 DTFANLVDVLSSEETILNLAKADISPKYIEINLLGCNMFSSYSISAEETYPGKLLKIK 720
 DB 661 DTFANLVDVLSSEETILNLAKADISPKYIEINLLGCNMFSSYSISAEETYPGKLLKIK 720

QY 721 DRVSELMPISQDSITVSANQVEYVINEEGKEIILDHSGKWINKEESIIKDISKEYISF 780
 DB 721 DRVSELMPISQDSITVSANQVEYVINEEGKEIILDHSGKWINKEESIIKDISKEYISF 780

QY 781 NPKENKIIYKSKYLHELSTLLOEIRNANSDDIDLEKVMLTECEINVASNIDRQIVEGR 840
 DB 781 NPKENKIIYKSKYLHELSTLLOEIRNANSDDIDLEKVMLTECEINVASNIDRQIVEGR 840

QY 841 IBEAKNLTSDSINYIKNEFKLIESDSLYDLKHONGLDSDHISFEDISTENGFRIRF 900
 DB 841 IBEAKNLTSDSINYIKNEFKLIESDSLYDLKHONGLDSDHISFEDISTENGFRIRF 900

QY 901 INKETGNSFIETEKEIFEYATHISKEISNKTIFDNVNGKLVKYNLDAAEVNTLN 960
 DB 901 INKETGNSFIETEKEIFEYATHISKEISNKTIFDNVNGKLVKYNLDAAEVNTLN 960

QY 961 SAFFIOSLIYNTTKESJNLVAMKVQVYAOQFSTGLNTITDASKVELVSTALDETID 1020
 DB 961 SAFFIOSLIYNTTKESJNLVAMKVQVYAOQFSTGLNTITDASKVELVSTALDETID 1020

RESULT 2
 Q9EXR0 PRELIMINARY; PRT: 2367 AA.

AC Q9EXR0;
 DT 01-MAR-2001 (TREMELrel. 16, Created)
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE Toxin B.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_taxid=1496;
 RN [1]
 RC STRAIN=8864;
 RA von Bichel-Streiber C.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RC STRAIN=8864;
 RA Kohl M.;
 RL "The pathogenicity locus of Clostridium difficile, strain 8864.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR PIR; A27636; A27636.
 DR GO; GO:0003743; F:translation initiation factor activity; IEA.
 DR GO; GO:0006413; P:translational initiation; IEA.
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR007577; Gly_transf_sug.
 DR InterPro; IPR001950; TIF_SUI1.
 DR Pfam; PF01473; CW_binding 1; 16.
 DR Pfam; PF04488; Gly_transf_sug; 1.
 DR PROSITE; PS01118; SUI1_1_1.
 SQ SEQUENCE 2367 AA; 269072 MW; BF7087C92E7389E7 CRC64;

Query Match 79.1%; Score 4105.5; DB 2; Length 2367;
 Best Local Similarity 77.4%; Pred. No. 8e-149;
 Matches 790; Conservative 114; Mismatches 116; Indels 1; Gaps 1;

QY 1 MNLVNAQLOKQVYKFRIOEDYVAIINALEEYHNMSESSVVEKYLKLDINNLTDNYL 60
 DB 1 MSLVNAQLOKQVYKFRIOEDYVAIINALEEYHNMSESSVVEKYLKLDINNLTDNYL 60

QY 61 NTYKSGRNKALKKPKFYLTMVELEKNSLTPVEKNLHF*WIGQINDTAINYNQWKD 120
 DB 61 NTYKSGRNKALKKPKFYLTMVELEKNSLTPVEKNLHF*WIGQINDTAINYNQWKD 120

QY 121 VNSDYFVKVYDSNAPLINTLKTIVESATNTLTSPRENLDPEFDYKFRMRMIY 180
 DB 121 VNSDYFVKVYDSNAPLINTLKTIVESATNTLTSPRENLDPEFDYKFRMRMIY 180

QY 181 DKQKHFIDYKSOIENPEFIIDNTIKTYLSNEYSKDLEALNKYIEESLNKITTANNNDI 240
 DB 181 DKQKHFIDYKSOIENPEFIIDNTIKTYLSNEYSKDLEALNKYIEESLNKITTANNNDI 240

QY 241 RNLEKFAEDLVRNLQELVERWNLAAASDIIRISMLKEDGGVLDVDILPGIOPDLFKS 300
 DB 241 RNLEKFAEDLVRNLQELVERWNLAAASDIIRISMLKEDGGVLDVDILPGIOPDLFKS 300

QY 301 INKPDSITNTSWEMIKLEALIMKYKEYIPGYTSKNPDMLEDEEVQSFESALSSKSDKSEI 359
 DB 301 INKPDSITNTSWEMIKLEALIMKYKEYIPGYTSKNPDMLEDEEVQSFESALSSKSDKSEI 359

QY 360 FLPLDDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINQIKRKYILNDNLNPSIN 419
 DB 360 FLPLDDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINQIKRKYILNDNLNPSIN 419

QY 420 EGTDFNTMTKIPSDKLASISNEDNMFMFKITNLYKVGFPADVRSTINLSGPGVVTGAYQ 479
 DB 420 EGTDFNTMTKIPSDKLASISNEDNMFMFKITNLYKVGFPADVRSTINLSGPGVVTGAYQ 479

QY 421 QGNDFTMTNMFESLGAIANEENISFIKISYLVGVFPYEPANTTITLSGPTIYAGAYK 480
 DB 421 QGNDFTMTNMFESLGAIANEENISFIKISYLVGVFPYEPANTTITLSGPTIYAGAYK 480

Qy 480 DLLMFKDNTNHLPELNPFPKTKISOLTEQBITSLWSFNQARAKSQEYKKGYP 539
Db 481 DLLTFKMSIDTSLSSLELNPFPKNTSQATEQKNSLWQNEERAKIQEYKKNYP 540
Qy 540 EGALGEDNDLDFACNTVLDKDYVSKKILSSMKTRNKEYIHYIVLOQDKISYEASCNLFS 599
Db 541 EGALGEDNDLDFACNTVLDKDYVSKKILSSMKTRNKEYIHYIVLOQDKISYEASCNLFS 600
Qy 600 KDPYSSILYQKNIEGSETATYVYVADAEIKEDIKYRIPYQISNKRNIKLTFIGHGKSEFN 659
Db 601 KNPYDSILFQKNIEDSEVAYVYVADAEIKEDIKYRIPYQISNKRNIKLTFIGHGKSEFN 660
Qy 660 TDTFANLVDLSLSEIETILNLAKADISPKYIIEINLLGCMNFSYSISABETYPGKLLKI 719
Db 661 TDTFANLVDLSLSEIETILNLAKADISPKYIIEINLLGCMNFSYSISABETYPGKLLKI 720
Qy 720 KDRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWKNEESIIKDISKEYIS 779
Db 721 KDRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWKNEESIIKDISKEYIS 780
Qy 780 FNPENKIIIVKSKYLHELSTLLOEIRNNSDDILEKKVMLTECEINVASNIDROIVEG 839
Db 781 FNPENKIIIVKSKYLHELSTLLOEIRNNSDDILEKKVMLTECEINVASNIDROIVEG 840
Qy 840 RIEBAKNTSDSINYIKNEFKLIESISDSLYDLKHQNGLDSDHFSIFEDISKTEGFRIR 899
Db 841 RIEBAKNTSDSINYIKNEFKLIESISDSLYDLKHQNGLDSDHFSIFEDISKTEGFRIR 900
Qy 900 FINKETGNSIFETETKEIFSEYATHISKEISNIDTIFDNGKLVKKNVLDAAHEVNTL 959
Db 901 FINKETGNSIFETETKEIFSEYATHISKEISNIDTIFDNGKLVKKNVLDAAHEVNTL 960
Qy 960 NSAFITQSLIEYNTTKESINLSVAMKVQVYAOLFSTGLNTITDASKVVELVSTALDETI 1019
Db 961 NSAFITQSLIEYNTTKESINLSVAMKVQVYAOLFSTGLNTITDASKVVELVSTALDETI 1020
Qy 1020 D 1020
Db 1021 D 1021

RESULT 3
Q9F931 PRELIMINARY; PRT: 2367 AA.
ID Q9F931
AC Q9F931;
DT 01-MAR-2001 (TremBrel. 16, Created)
DT 01-MAR-2001 (TremBrel. 16, Last sequence update)
DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
DE Cytotoxin B.
GN TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_taxID=1496;
RN [1]_taxID=1496;
RC STRAIN=5340;
RX MEDLINE=20448897; PubMed=10992443;
RA Sambol S.P., Merrigan M.M., Lively D., Gerdling D.N., Johnson S.;
RT "Toxin gene analysis of a variant strain of clostridium difficile that
causes human clinical disease";
RL Infect. Immun. 68:5480-5487(2000).
DR EMBL; AF217292; AAG18011.1; -
DR PIR; A27636; A27636.
DR GO; GO:0003743; P:translational initiation factor activity; IEA.
DR GO; GO:0006413; P:translational initiation; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR005777; Gly_transf_sug.
DR InterPro; IPR001950; TIF_SUI1.
DR Pfam; PF01473; CW_binding_1; 17.
DR Pfam; PF04488; Gly_transf_sug; 1.
DR PROSITE; PS01118; SUI1_1; 1.

SQ SEQUENCE 2367 AA; 269337 MW; D5BE715ESBD41E2F CRC64;
Query Match 78.9%; Score 4096.5; DB 2; Length 2367;
Best Local Similarity 77.0%; Pred. No. 1.8e-148;
Matches 786; Conservative 117; Mismatches 117; Indels 1; Gaps 1;
Qy 1 MNLVKAQLOQVAVVVKRIQEDVEVAILNALVEYHNSSESVKYLKLDINLNTDYL 60
Db 1 MSLVNRKQLEKQANRFRVQDEYVAILDALVEYHNSSESVKYLKLDINLNTDYL 60
Qy 61 NTKKSGRNKALKFKKYLWTEVLELKNLSLTPVEKQLHFIWGGQINDTAINYNQWD 120
Db 61 DTYKSGRNKALKFKKYLWTEVLELKNLSLTPVEKQLHFIWGGQINDTAINYNQWD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLTSPRENINDEPDKYFRKMEIY 180
Db 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLTSPRENINDEPDKYFRKMEIY 180
Qy 181 DKQKHFDYKSOEENPEFIIDNIIKTYLSNEYSKOLEALNKYIEESLNKITANGNDI 240
Db 181 DKQKHFDYKSOEENPEFIIDNIIKTYLSNEYSKOLEALNKYIEESLNKITANGNDI 240
Qy 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRLSMLEKDGGVYLDVILFGIOPDLFKS 300
Db 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRLSMLEKDGGVYLDVILFGIOPDLFKS 300
Qy 301 INKPDISI-TNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSKSDSEI 359
Db 301 INKPDISVTAVDEEMQLEAIMKHYKEYIPGYTSKNFMDLDEEVQSFESALSKSDSEI 360
Qy 360 FLPLDDIKVPLEVIAFANNVINOALISKDSYCSDLVINQIRKYLINDNLNPSIN 419
Db 361 FLPLDDIEVPLEVIAFAGKSIINOALISAKDSYCSDLLIKQIRKYLINDNLNPSIN 420
Qy 420 EGTDPNTTKI-FSKLASISNEDNMWMIKITNLYKVPAPDVRSTINLSPGVYTGAYQ 479
Db 421 QGNDFTTMMNFGESLGAIBENISFIATKIGSVLRGVFPYEAINTVTLSTGPTIYAGAYK 480
Qy 480 DLLMEKDNSTNHLPELNPFPKTKISOLTEQBITSLWSFNQARAKSQEYKKGYP 539
Db 481 DLLTFKMSIDTSLSSLELNPFPKNTSQATEQKNSLWQNEERAKIQEYKKNYP 540
Qy 540 EGALGEDNDLDFACNTVLDKDYVSKKILSSMKTRNKEYIHYIVLOQDKISYEASCNLFS 599
Db 541 EGALGEDNDLDFACNTVLDKDYVSKKILSSMKTRNKEYIHYIVLOQDKISYEASCNLFS 600
Qy 600 KDPYSSILYQKNIEGSETATYVYVADAEIKEDIKYRIPYQISNKRNIKLTFIGHGKSEFN 659
Db 601 KNPYDSILFQKNIEDSEVAYVYVADAEIKEDIKYRIPYQISNKRNIKLTFIGHGKSEFN 660
Qy 660 TDTFANLVDLSLSEIETILNLAKADISPKYIIEINLLGCMNFSYSISABETYPGKLLKI 719
Db 661 TDTFANLVDLSLSEIETILNLAKADISPKYIIEINLLGCMNFSYSISABETYPGKLLKI 720
Qy 720 KDRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWKNEESIIKDISKEYIS 779
Db 721 KDRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWKNEESIIKDISKEYIS 780
Qy 780 FNPENKIIIVKSKYLHELSTLLOEIRNNSDDILEKKVMLTECEINVASNIDROIVEG 839
Db 781 FNPENKIIIVKSKYLHELSTLLOEIRNNSDDILEKKVMLTECEINVASNIDROIVEG 840
Qy 840 RIEBAKNTSDSINYIKNEFKLIESISDSLYDLKHQNGLDSDHFSIFEDISKTEGFRIR 899
Db 841 RIEBAKNTSDSINYIKNEFKLIESISDSLYDLKHQNGLDSDHFSIFEDISKTEGFRIR 900
Qy 900 FINKETGNSIFETETKEIFSEYATHISKEISNIDTIFDNGKLVKKNVLDAAHEVNTL 959
Db 901 FINKETGNSIFETETKEIFSEYATHISKEISNIDTIFDNGKLVKKNVLDAAHEVNTL 960
Qy 960 NSAFITQSLIEYNTTKESINLSVAMKVQVYAOLFSTGLNTITDASKVVELVSTALDETI 1019
Db 961 NSAFITQSLIEYNTTKESINLSVAMKVQVYAOLFSTGLNTITDASKVVELVSTALDETI 1020

```

QY 1020 D 1020
Db 1021 D 1021

RESULT 4
Q46034 PRELIMINARY; PRT; 2367 AA.
AC Q46034
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Toxin B.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate 1470;
RX MEDLINE=96079281; PubMed=7494480;
RA von Eichel-Streiber C., Meyer zu Heringdorf D., Habermann E.,
RA Saringen S.;
RT "Closing in on the toxic domain through analysis of a variant
RT Clostridium difficile cytotoxin B.";
RL Mol. Microbiol. 17:313-321(1995).
DR EMBL; Z23277; CAA80815.1; -.
DR PIR; A27636; A27636.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR GO; GO:0006413; P:translational initiation; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR001950; TIF SUI1.
DR Pfam; PF01473; CW binding 1; 17.
DR Pfam; PF04488; Gly transf_sug; 1.
DR PROSITE; PS01118; SUI1_1; 1.
SQ SEQUENCE 2367 AA; 269186 MW; BP9823DAE70427F3 CRC64;

Query Match 78.8%; Score 4088.5; DB 2; Length 2367;
Best Local Similarity 77.1%; Pred. No. 3.6e-148;
Matches 787; Conservative 114; Mismatches 119; Indels 1; Gaps 1;

QY 1 MSLVKAQOKVYKFRQEDBYVAILNALBEYHNMSESSVVEKYLKLDINNTDYL 60
Db 1 MSLVNRKQLEKQVAVRQEDBYVAILDALBEYHNMSESSVVEKYLKLDINSLTDYI 60
QY 61 NTYKSGRNKALKKPKKYLTMVEVLELKNLSLTPVEKNLHFWIGGOINDTAINYNQWD 120
Db 61 DTYKSGRNKALKKPKKYLTVILELKNLSLTPVEKNLHFWIGGOINDTAINYNQWD 120
QY 121 VNSDYFVKVYDSNAFLINTLTKTIVESATNTLSEFRENLDNPEFDYKFKRMEIY 180
Db 121 VNSDYNVNFYDSNAFLINTLTKTIVESATNTLSEFRENLDNPEFHTAFKRMQIY 180
QY 181 DKQKHFIDYKSGIEENPEFIIDNIKTYSNYSKDLKALNKYIEESLNKITANNNDI 240
Db 181 DKQQNFYNYKAKQENPDIIIDIVKTYLSNYSKDIIDELNAYIEESLNKVTENSGNDV 240
QY 241 RNLEKFADEDLVNLQELVERVNLAAASDIIRISMLKEDGGVYLDVILPGIQDPLKS 300
Db 241 RNPEEFKTEGVNLYEQESVERVNLGASDILRVAILKNIGVYLDVMDLPGIHPDLFXD 300
QY 301 INKPSDI-TNWSNEMIKLAIWKYKEYIPGYTSKNFDMLEDEVSFESALSSDKSEI 359
Db 301 INKPSVKTAVDWEENQLEAIWKYKEYIPGYTSKNFDMLEDEVSFESVLSKSDKSEI 360
QY 360 FLPLDDIKVSPLEVKIAPANNVINOALISLSDYSGLVINQIKNRYKILNDNLNPSIN 419
Db 361 FLPLGLIEVSPLEVKIAPAKGSIINGALISAKDSYCSDLIIKIQIKNRYKILNDTGPIS 420
QY 420 EGTDFNTWKIFSDKLIASINEDNMFMKIKITNLYKVGAPDVRSTINLSGPGVYTGAYQ 479

```

Query Match 41.6%; Score 2156.5; DB 2; Length 554;

```

RESULT 5
Q93139 PRELIMINARY; PRT; 554 AA.
AC Q93139
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE TCD-B-C34 Cluster1-2 (Fragment).
GN TCD-B-C34.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C34;
RX MEDLINE=21323121; PubMed=11430410;
RA Mehlig M., Moos M., Braun V., Kalt B., Mahony D.E.,
RA von Eichel-Streiber C.;
RT "Variant toxin B and a functional toxin A produced by Clostridium
RT difficile C34.";
RL FEMS Microbiol. Lett. 198:171-176(2001).
DR EMBL; AJ294944; CAC41640.1; -.
DR InterPro; IPR007577; Gly transf_sug.
DR Pfam; PF04488; Gly transf_sug; 1.
FT NON TER 554
SQ SEQUENCE 554 AA; 63910 MW; 9285512AC1B1D87A CRC64;

```



```
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Toxin A (Fragment).
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C34;
RX MEDLINE=20402122; PubMed=10931284;
RA Braun V., Mehlig M., Moos M., Rupnik M., Kalt B., Mahony D.,
RA von Eichel-Streiber C.;
RT "A chimeric ribozyme in Clostridium difficile combines features of
RT group I introns and insertion elements.";
RL Mol. Microbiol. 36:1447-1459(2000).
DR EMBL; AJ131844; CAC03681.1; -.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF04488; Gly_transf_sug; 1.
FT NON_TPR 553 -553
SQ SEQUENCE 553 AA; 63913 MW; DDE4551A6D8C3B25 CRC64;

Query Match      28.4%; Score 1474; DB 2; Length 553;
Best Local Similarity 51.7%; Pred. No. 5,1e-49;
Matches 287; Conservative 108; Mismatches 158; Indels 2; Gaps 2;

QY 1 MNLVWKAQLQKVVYKFRIOEDVEVAILNALAEYHNMSESSVVEKYLKLDINLTDNYL 60
DB 1 MSSISKELIKLAY-SVRPRENEYKITLNDYENKLTNNNNENKYLQKLKLNESIDVFM 59

QY 61 NTYKSGRNKALKKFKYLTMEVLELKNLSLTPVEKNLHFIWIGGOINDTAINYINQWD 120
DB 60 NKNYNSNRNALSNLKKDILKEVILKNSNTSPVEKNLHFWIGGEVSDIAEYIKQWAD 119

QY 121 VNSDYTVKVPYDSNAFLINTLTKTIVESATNTLESFRENLPDPEDYKFKRMEIY 190
DB 120 INAEYVWLWYDSAPLVNTLKKALVESSSTTEALQLLEBEIQNPQDM-KFYKRWEIFY 178

QY 181 DKQGHFDYYSQIEENPEFIIDNIKTLSNEYSKDLEALNKYIEESLNKTANNNDI 240
DB 179 DRQKRFINYKQINKPTVPTDIDIKSHLVSEYNRDETLLSEYRTNSLRKINSNHGIDI 238

QY 241 RLNEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVLDILPGIQPLFKS 300
DB 239 RANSLFTEQELNLYSQELNLRGTLAAASDIVALLAKNFGVYLDVDMPLGHSDFKT 298

QY 301 INKPDSTINTSWEMIKLEAMKYKEYIPGYTSKNFMDLDEVORSFESALSSKSKSEIF 360
DB 299 IPRSSSGLDRWEMIKLEAMKYKYINNTYSENFDKLDQQLKONFKLIESKSEIF 358

QY 361 LPLDDIKVSPLEVKIAPANNVINOALISLKDSYCSDLVINOIKRYKILNDNLPSINE 420
DB 359 SKLENLVSDLEIKIAPALGSVINOALISKQGSYLTNLVIEQVKRYQFLNQHLPAPIS 418

QY 421 GTDENTMTKFSDKLASISNEDNMFMKITYNLKVGFPAPDVRSTINLSGPGYTTAYQD 480
DB 419 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPLVQGVMPPEARSTISLSGPGYASAYD 478

QY 481 LLFMKDNSTNHLPELRNPEPKTKISOLTEQIITSLWSNOARAKSOFERYKKGYFE 540
DB 479 FINQENTIEKTLKASDLIEFKFPENNLSQITEQINSLSWFSQASAKYQIERYVDYTG 538

QY 541 GALGEDNLDFAQNT 555
DB 539 GSPSGDNGVDNFKNT 553

RESULT 10
Q46149 ID Q46149 PRELIMINARY; PRT: 2178 AA.
AC Q46149; Q46147; Q46148;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alpha-toxin.
OS Clostridium novyi.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19402;
RX MEDLINE=95342160; PubMed=7616958;
RA Hofmann F., Herrmann A., Habermann E., von Eichel-Streiber C.;
RT "Sequencing and analysis of the gene encoding the alpha-toxin of
RT Clostridium novyi proves its homology to toxins A and B of Clostridium
RT difficile.";
RL Mol. Gen. Genet. 247:670-679(1995).
RN [2]
RP SEQUENCE OF 1204-2178 FROM N.A.
RC STRAIN=ATCC19402;
RA Hofmann F., Habermann E., von Eichel-Streiber C.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z48636; CAA88565.1; -.
DR EMBL; Z23280; CAA80818.1; -.
DR EMBL; Z23281; CAA80819.1; -.
DR PIR; S55805; S55805.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF04473; CW binding_1; 10_
DR Pfam; PF04486; Gly_transf_sug; 1.
SQ SEQUENCE 2178 AA; 250134 MW; 9B0ADCE031C4A75A CRC64;

Query Match      27.7%; Score 1435.5; DB 2; Length 2178;
Best Local Similarity 32.8%; Pred. No. 6.7e-47;
Matches 343; Conservative 212; Mismatches 435; Indels 55; Gaps 19;

QY 3 LVNKAQLQKVVYKFRIOEDVEVAILNALAEYHNMSESSVVEKYLKLDINLTDNYLN 61
DB 2 LITREQLMKIASIPLKRKEPEYNLILDALENFNRDIEGTSVKETYSKLSKLNLVDNYQT 61

QY 62 TYKSGRNKALKKFKYLTMEVLELKNLSLTP--VEKNLHFIWIGGOINDTAINYINQWK 119
DB 62 KYPSSGRNALLENFRDLSYSELRELIKNSRTSTTASKNLSFIWIGGPIQSQSLYNNWK 121

QY 120 DVNSDYTVKVPYDSNAFLINTLTKTIVESATNTLESFRENLPDPEDYKFKRMEI 179
DB 122 MFNDYINRLFYDKNSLLVNTLTKTALQESSKVIIEQNSNLDGTGHNKFPYSDRMKLI 181

QY 180 YDKQGHFDYYSQIEENPEFIIDNIKTLSNEYSKDLEALNKYIEESLNKTANNND 239
DB 182 YRYKELKMLYENKQNS---VDDIINFLSNYFKYDIGKLNQKNNNNKMLAIGATD 238

QY 240 IRNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVLDILPGIQPLFK 299
DB 239 I-NTENILTNLKLKSYVQELIQTNLAAASDILRILKYGKGYVCDLFLPGVNLSLFN 297

QY 300 SINKPDSINTSWEMIKLEAMKYKEYIPGYTSKNFMDLDEVORSFESALSSKSKSEI 359
DB 298 DISKPNGMSNYWEAAIFEAIANEKKLMNNYPYKYMGOVPESEIKERILSFVRN-HDINDL 356

QY 360 FLPLDDIKVSPLEV-----KIAFANNVINOALISLKDSYCSDLVINOIKRYKILNDNL 414
DB 357 ILPLGDIKLSQELLSLKLKATGKTFSNAFIISNNDLSLTNNLSOLENRYEILNSII 416

QY 415 NP--SINEGTD--FNNTMTKFSDKLASISNEDNMFMKITYNLKVGFPAPDVRSTINLSG 470
DB 417 QEKFKICETYDSYNSVSELVLETPKNLSMDGSSFYQIIGLYSSGFKPEVNSTVPSFG 476

QY 471 PGVYTGAYQDILMFKN-----STNHLPELRNPEPKTKISOLTEQIITSLWSNPQ 524
DB 477 PNIYSATCDTVHFIKNTFDMLSNQOEIPEAS--NNLYFSK-----THDEFKSKSWLRS 529

QY 525 ARAKSQPEYKGYEGEGALGEDNLDPAQNTVLDKDYVSKK--LSSMKTKNKYIHVIV 582
DB 530 NIAEKEFOKLITYIGRTLNVEDGLNFKNKRVTTSLLKVIIBEVNSTKIYENYDNLNML 589
```


QY 583 QLOQDKISYEASCNLFSDPYSSILYQKNIEG-SETAYVYVADAIEIKDKRIPYQIS 641
 Db 590 QIQGDDISYASVNVFGKPKNSIL-IOGVDDFANVFENGIVQSDNINNLSRFN 645
 QY 642 NKRNLKLFTHGKSEFNTDFANLVDLSSEI-----ETILAKADISPKYIEINLL 696
 Db 646 DIKKIKLTLIGHGVNFPKLPFGGKTVDLVNTIIPKLOHLLREGVILKVKYKLNIL 705
 QY 697 GCMFYSISABETVPKLLIKIKDRVSELP-SISQSDISVSANOYEVIRINEGKEBIL 755
 Db 706 GCMFPTPKVDINSTTVGLFNKIS-RDLOPKGFSKQKLSANVKAIRINRGKREVL 762
 QY 756 DSHGKWKNEBESIIKDISKEYISPNKENIIVKSKYLHELSTLLOQIRNANSSDIDL 815
 Db 763 DIFGKWSNTDIIAQISNKYVYVNEVENTLSARVEQKLVKAFKDI-----NSLIQT 817
 QY 816 EKKVWLTECEINVAGNIDRQIVEGRIEAKNLTSDSI-----NYIKNEFKLIES:SDSLYD 871
 Db 818 TNNQELKQSLVNTYADLTITLYSELKEDIPELDNIQIKERIILNEISRLHDFSNILD 877
 QY 872 LKHQGLDSDHSPISFEDISKTENGPRIRPINKETGNSIFETEKEIFSEYATHISKEISN 931
 Db 878 FYQKNISNMILFDSIIKEKDYNVKLANKITGETSVIKTYSDSLWNFTNKYKIVDD 937
 QY 932 IKDTIFDVTNGKLVKWNLDAAHEVNTLSAFQISLIEYNTTKESLNSLVAMKVQVYA 991
 Db 938 IRGIIIVKIDNGEFKADPEIQNSPLNSAMLMQLLDIDYPTBELTNNMTSLUKVQAYA 997
 QY 992 QLPSTGLNITDASKVVELVSTALD 1016
 Db 998 QIFQLSIGAIQATEIVTIISDALN 1022

RESULT 11
 Q9ZGR4
 ID Q9ZGR4 PRELIMINARY; PRT; 3169 AA.
 AC Q9ZGR4;
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Putative cytotoxin.
 GN L7095.
 OS Escherichia coli O157:H7.
 OC Plasmid pO157.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EDL933;
 RX MEDLINE=98391744; PubMed=9722640;
 RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
 RA Blattner P.R.;
 RT "The complete DNA sequence and analysis of the large virulence plasmid
 of Escherichia coli O157:H7";
 RL Nucleic Acids Res. 26:4196-4204 (1998).
 DR EMBL; AF074613; AAC70163.1; .
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR InterPro; IPR007577; Gly_transf_sug.
 DR InterPro; IPR006473; Peptidase_C58_yt.
 DR Pfam; PF04488; Gly_transf_sug; 1.
 DR TIGRFAMs; TIGR01586; ycpT_cys_prot; 1.
 KW Plasmid.
 SQ SEQUENCE 3169 AA; 361971 MW; C2CDB48A72BD5FD1 CRC64;

Query Match 8.1%; Score 422.5; DB 2; Length 3169;
 Best Local Similarity 21.0%; Pred. No. 4.9e-08;
 Matches 258; Conservative 197; Mismatches 450; Indels 323; Gaps 54;

QY 14 YVFR1---QEDVAILNAL--BEYHNSESSVWEKYLKDKDINNLTNYLNTYKSGR 68
 Db 160 YIKIRTRGAEDQTTTITQSLIINELINGVDRNTI-PPQKISLNDIHSYENMQIKSR 218

QY 69 N--KALKKFEYLTWELVELKNNS-----LTPVEKNLHFIWIGQ 106
 Db 219 KGEILVKQGEILLSSLNDKNGKQSDNASKIINLLGIEYQSHKVDIEPTTHAVWAGA 278
 QY 107 INDTAIYNQKVDNSDYTVKVFVDSNAF-----LINTLKKT----- 144
 Db 279 PPDNTFSYITAFNTYKYDTYLLWIDPNAFGAAPSGILKNIAMNYAIVMRRTNPHLAE 338
 QY 145 -----IVESATNNTLE--SPRENLDPEPY-----NKFYRKMEIIEYDQKHFI 187
 Db 339 ENNEVLK:QNTQNETIEFEKTERELKELENYKSLTSEKFNVFVLESIGWQDNVYF 398
 QY 188 DYYKSGIEENPEFI-----IDNIIK--TYLSNEYSKDLEALNKYIEESLUNKITANNGN-- 238
 Db 399 TYCISNGISNTDDISRLDFTLVKLSPVQNDFKSTVEKNKRDIDLLKNTISQKGF 458
 QY 239 ---DIRNLKPADEDLVRLYNQELVERWLAASILIRISMLKEDGGVLDVLDLPGIOP 295
 Db 459 QLRDINTLESFKKPDQYFFYOQEMLRWYAAASQVRINILKEYGGIYTDITLPAYSO 518
 QY 296 DLPFSINKPDSITNTSWEMIKLEAMKYK--EVIQ--YTSKNFMDLDEVQSFESALS 351
 Db 519 KVSQIINE-KSDDKRFEDLKLRIISELSILIKGEKYSIKH-DGLDDETLNQLNLI- 575
 QY 352 SKGDKSEIFLPLDDIKVSPLEVKIAFANNSVI-----NQALISLKSQSYC 395
 Db 576 SEIEK-----LTIDY--FKPVEIKVYEDTFKIPKQKWTENTWIRGNPNFLTHGSKC 630
 QY 396 SDLVINQIKNRYKILNDNLNPSINEGTFTNTWKIPSDKLASISNE-----DNM 444
 Db 631 IDFILSQKKQYLEL-QRIRDNISYNNLFVT-----EDLKSINNVAIGGIPAKKYLEHG 684
 QY 445 MFMIKITNLYKGFADVRSTINLSCP-----471
 Db 685 LF-----SEYRQDGTIPYVYSTLNISGPFDMRMQMKYKSLGRIEYVHTKONKLSVNL 740
 QY 472 GVTGAYQDILMF---KDNSTNIHLLEPE-----LRNPEFPKTKISQLT---EOEITS 518
 Db 741 GTVASSNKNKNSFWLNPVSVGINDITPDESSWAVRNNDINKILKEKINCHVPEKLPTS 800
 QY 519 LWSFNQARAKSQPEYKKGFGALGEDDNLDPQNTVLDKQYVSKILSSMKTRKNEYI 578
 Db 801 LY-----YEDRSRPFQW---DNKSIKHVTEINKOLI--KDINLLTSSNIDV 844
 QY 579 HYIVQOGDKISYEASCNLFSDPYSSILYQKNIEGSETAYVYVADAIEIKDKRIPY 638
 Db 845 KLLIKL--DRELYAISKI-----DNPLALRSIRTLQQLANYVTSNTPEPENTINFIY 896
 QY 639 QISNKN-----IKLTFIGHKSEFNTDFANLDVDS-----LSSEIET 677
 Db 897 DFYRKQDOLLSA:KL-----FSRNDADTKIIVWYNSVMEKNVLEFREVISC 942
 QY 678 ILNLKADISPKYTEINLLGCMNFSYSISABETVPKLLIKIKDRVSELMPSISQDSI-- 735
 Db 943 VLRSKKVD--SVINEN-----KNLSKEDA--GALRDYAKLKKMELFSLMDDDGYKK 990
 QY 736 TVSANQYEVIRINEEGREILDHSGKMNKBEESIKDISKEYISFNP-----K 793
 Db 991 IITTNAY--IKERDKL-----SGIIVNIENSIISGHESPDIIIRSNQHEWGLSTVEQFK 1042
 QY 784 ENKIIIVKSKYLHELSTLLOQIRNANSSDIDLEKKVWLFECEINVAS-----NIDQIVE 838
 Db 1043 KFEYFVKSE-LSSAKSIFDDIKNK-YITDPTKRNVLVHOLDSDIKERIAFLDISHYAYP 1100
 QY 839 GRIBEAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHQNGL 878
 Db 1101 GSLLEKIQLS-----GYVFSIDNIIAEYVLASYGVSHGVYVPAPSKLLELRHTK 1156
 QY 879 DDSHFIFEDISKTENGFRIRFINKETGNSI---FIETEKEIFSEYATHISKEISNIKT 935
 Db 1157 SNSEWI--EKITP-----YVVDIISDNVSVLRPLSEEQKILNDIKLEISKSVS---EQ 1207

```
Qy 936 IFDINVNGKLVKVNLDAAHEVNTLSNAFFQSIIEYVNTK-----ESLSNLSV----- 993
Db 1208 YFMKLTQKSSVIGIKYVDFDRVNEFLSLPLNQNLTPFMYRYFEMLYDIHIGIEN 1267
Qy 984 -AMKVQVYQALFSTGLNTITDASKVVEL 1010
Db 1268 KANREFIYSKSSLNLDLFLINDERVLNL 1295

RESULT 12
ID 082916 PRELIMINARY; PRT; 3169 AA.
AC 082916;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Toxin B.
GN TOXB.
OS Escherichia coli O157:H7.
OG Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7;
RX MEDLINE=96290540; PubMed=9628576;
RA Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,
RA Yutsudo H.C., Kubota Y., Yamauchi Y., Iida T., Yamamoto K., Honda T.,
RA Han C.G., Ohtsubo E., Kasamatsu M., Hayashi T., Kuhara S.,
RA Shinagawa H.;
RT "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an
RT enterohemorrhagic Escherichia coli O157:H7 derived from Sakai
RT outbreak.";
RL DNA Res. 5:1-9(1998).
DR EMBL; AB011549; BAA31815.1; -.
DR PIR; T00296; T00296.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR007577; Gly trans aug.
DR Pfam; PF04488; Gly trans aug; 1.
DR TIGRFAMs; TIGR01586; yopT_Cys_prot; 1.
KW Plasmid.
SQ
SEQUENCE 3169 AA; 361943 MW; D9BED726A62DDBFF CRC64;

Query Match 8.1%; Score 419.5; DB 2; Length 3169;
Best Local Similarity 20.8%; Pred. No. 6.4e-08;
Matches 256; Conservative 202; Mismatches 446; Indels 325; Gaps 55;

Qy 14 YVKFRI---QDEYVAILNAL--BEYHNMSESVYKYLKLDINNLTDNYLNTYKSGR 68
Db 150 YIKIRTRGAEQTITTSQSLINELLNGVDRNTI-PFQKISELNDIHSVENMQIKNSR 218
Qy 69 -----NKAALKPKE----YLTMEVLELKNLSLTPVEKNLHFTWIGG 105
Db 219 KGIEILVKGELLSSLNIVNKGKQLSDNASKIINLIGIEYQSHK-VDIEPFIHAVVAG 277
Qy 106 QINDTAINYINQWDVNSDYTVKYFYDSNAF-----LINTLTKT----- 144
Db 278 APPDNTFTYITAFNLTYKDYTYLLWDIPNAGAKFGILKNIAMNVAIMELRARTNPHLA 337
Qy 145 -----IVESATNNTLE--SFRENLDPEFDY-----NKFYKRMELIYDKQKH 186
Db 338 EEMNEVILKIQNIQNETIEFKETERLEKLENRYKSLTSEKFEKNVFFLESIGMQDNY 397
Qy 187 IDYKSKQIEENPEFI-----DNLIK--TYLSNYSKDLKALNKYIESLNKNTANNGN- 238
Db 398 FTYCISNGISNTDDISRLDFTLVNLKLSPEQNDPFKSTVEKNKRDILLKNTISQKFGDR 457
Qy 239 -----DIRNLEKFADEDVRLVNLQELVERWNLAASDLIRISMLXEDGGVYLDVILG 294
Db 458 FQLRDINTLESFKPKPDYFFYQEMLLRWNYAAASDQVRINILKYEYGGIYDITDILPAYS 517
```

```
Qy 295 PDLFKSINKPSITNTSWEMIKLEAIMKYK--EYIPG--YTSKXFDMLDEVQVSFSAL 350
Db 518 DKVQSQLINE-KSDOKRFFEDLKLRIISSESILSIKGEKYSIKH-DGLDETTLNQLNAIL 575
Qy 351 SSKSDKSEIFLPLDDIKVSPLEVKIAFANNSVI-----NQALISLSDSY 394
Db 576 -SEIEK-----LTIDDY-FKPVETKVRDTPKIFKRYOKWIENTWNIRGNMFMTHKSGK 629
Qy 395 CSDLVINQIKNRYKILNDNLNPSINEGTDFNTTKIFSDKLASISNE-----DN 443
Db 630 CIDFILSGCKQKYLEL-QRIRDNISYNNLFYTT-----EDLKSNNVAIGGIPAKKYLEH 683
Qy 444 MMFMIKITNYLKVGFAPDVRSTINLSGP----- 471
Db 684 GLF-----SEYRQDGTIPYVWSTLNTSGPDMRMQMKYKSLGRIGEVIHKDNKLSVNF 739
Qy 472 -GVYTGAYODLLMF---KDNSTNHLEPE-----LRNFEFPKTKISQLT---EQBIT 517
Db 740 LGVYASSNKDKNSFNWLPVSVGINDTIPDDESWAVANNNDINKILFEKINCHVPEKLPT 799
Qy 518 SLWSFNQARAKSQPEEYKGFEGALGDDNLDFAQNTVLDKVYVSKILSSMKTRNKEY 577
Db 800 SLY-----YEIDRSRFFQW---DNKSIKHVTEINKDLI--KDIINLLTSSNID 843
Qy 578 IHVIVLOQDKISYEASCNLFSDPYSSILYQKNIEGSETAYYYVADAIEIKDKYRIP 637
Db 844 VKLLIKL--DRELYAISKI-----DNPLALRSITLQLQLANVYTSNTPEPTINTFI 895
Qy 638 VOISNKRN-----IKLFTIGHGKSEPTDFTANLDVDS-----LSSIE 676
Db 896 YDFYRKQDQLLSAIKL-----FSRNDADTKIIVWYNSVMEKNVFLREVIS 941
Qy 677 TILNLAKADISPKYIEINLLGNMPSYISABEYYPGKLLKIKDRVSELMPSISQDSI- 735
Db 942 CVLRSEKVD---SYINEN-----KKNLSKEDA--GALRDYAKLKMKELFMSLDDDGK 989
Qy 736 -TVSANQYEVRIINBEGREILDHSGKWINKESIIKDISKEYISFNP----- 782
Db 990 KIITINAY--IKERDKL-----SGIYINENSIISGHESFDIIRSNQHEWGLSTVQEP 1041
Qy 783 KENKIIVKSKYLHELSTLLOEIRNANSDDIDLEKKVMLTECEINVAS-----NIDROI 837
Db 1042 KKFEFVVKSE-LSSAKSIFDDIKNK-VITDPETKENVLYHQDSDIKERIAFLDISHVAY 1099
Qy 838 EGRIEAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHONG 877
Db 1100 PGSILKELQLS---GYVFSIDNIIAEYLLASGVSGHYSHGVYVPAPSDKLELLRHT 1155
Qy 878 LDDSHFISFEDISKTEGFRIRFINKETGNSI---FIETKEIFSEYATHISKETSNIKD 934
Db 1156 KNSSEWI--EKITP---YVYDILSDNVSNVLRPPLSBEQKILNDIKLEISKVS--- 1206
Qy 935 TIFDINVNGKLVKVNLDAAHEVNTLSNAFFQSIIEYVNTK-----ESLSNLSV--- 983
Db 1207 QVFMKLTQKSSVIGIKYVDFDRVNEFLSLPLNQNLTPFMYRYFEMLYDIHIGIIE 1266
Qy 984 --AMKVQVYQALFSTGLNTITDASKVVEL 1010
Db 1267 KANREFIYSKSSLNLDLFLINDERVLNL 1295

RESULT 13
Q8L1L9
ID Q8L1L9 PRELIMINARY; PRT; 3223 AA.
AC Q8L1L9;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Bfal-Lymphostatin-like protein.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN-493/89;
RA Janka A.;
RT "Identification and Distribution of the Enterohemorrhagic Escherichia
RT coli Factor for Adherence (efal) Gene in Sorbitol Fermenting
RT Escherichia coli O157:H-";
RL Thesis (2002). Department of Institut fuer Hygiene,
RL University of Muenster, Muenster, Germany.
DR EMBL; AJ459584; CAD30848.1; -.
DR PIR; D91111; D91111.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001917; AminoTrans II.
DR InterPro; IPR007577; Gly transf sug.
DR InterPro; IPR006473; Peptidase C58 yt.
DR InterPro; IPR00169; SHprot acsite.
DR Pfam; PF04488; Gly transf sug; 1.
DR TIGRfam; TIGR01586; yopi_cys_prot; 1.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
DR PROSITE; PS00639; THIOLEPROTEASE_HIS; 1.
SQ SEQUENCE 3223 AA; 365885 MW; 1A65A380A7C6E6ED CRC64;

Query Match 7.5%; Score 391; DB 2; Length 3223;
Best Local Similarity 19.9%; Pred. No. 7.9e-07;
Matches 255; Conservative 205; Mismatches 430; Indels 392; Gaps 55;

QY 2 NLVNAQLQKMYVVFRIQEDBYVALNALAEYHNMSESS-----VVEKYLK-LKQINN 54
DB 226 NFVQKLSLELTKY---ANLEIKRIMGEYNLLPDKNSENGKLLQKQADLLKIME 282
QY 55 LTDNVLNTYKSGRNKALKKFKXEYLTMEYLEKN---NSLTPVEKNLHFIWIGQINDTA 111
DB 283 DISVTENTFKN-----TEMAIADIKREYVYSHTVDIEXNIIHAIWVAGSPPEI 329
QY 112 INYINQKDVNSDYTVKYFDYDANFLNTLTKTIYESATNTLESFRENLDNDFEDYK 171
DB 330 SDYIKTFKTYEFTYLVWDEKAFGAAKFTSVLKQIAFLACRTIQOQTQKNIDFINL 389
QY 172 Y---PKR-----MEIYDQKHFDY-- 189
DB 390 YNEIRKKNYNSGGOEYVNLKRELYATYQKISTPLKMFNSFFLENMVKLQDNFFNYCI 449
QY 190 YKSOEENPEFIID---NIITKYLNE---YSKOLEALNKYIEESL-----NKIT 233
DB 450 VKGVTINDELINYLKVIK-LSDDDIGNYQKTIINDNKQVKKLILDLQKQFGENRIS 507
QY 234 ANNGNDIRNLKPADEDLVRLYNQELVERNWLAAASDILRISMLKEDGGVYLDVDILPGI 293
DB 508 IKDVNSLTSLSKSENN---HNYQTEMLLRWNPAAADLLRMVYLKEHGGIYTDMDMPAY 564
QY 294 QPDLFKSINKPDSITNTSHEMILK-----EAMKYKEYIPGYTSKNFMDLD-EEVQSF 347
DB 565 SKQVIFKIMQTNQNGDRFLDLKLRRAISDGLVRY-----VNNQNIDEVYNEYI- 613
QY 348 SALSSKSDKSEIFLPDDDKVSPLEVKIAPANSV-----INQ 385
DB 614 ----SDADKNIIKILTEISKUPEDSIFTKILTRIPDTMPILRHYHLWPGWNRIGLNG 669
QY 386 ALISLSDKSDYCDLVI---NQIKNRYKILNDNLPNSINECTDFTNTWKIPFSDKLASIS 442
DB 670 FMLS HKGSEVIDAVIAGQVAYRELARIRDNHSEIY---FKQT-----DELSLPD 720
QY 443 NMFMFKITNYLKVQ-----FAPDVRSTINLSGP----- 471
DB 721 KIGGIL-VKKYLSGSLFSKFRQDTIIPALSTLIQISGPDLIQKMLQFFRSRGLVEBFI 779
QY 472 -----GVY-----TGAYQ-----DLMFKONS 488
DB 780 NERKLSDKAVIGYKTKTGSKYDNLTPADESTWICGKRCVDFLFDKQVS 839

```

```

QY 489 TNHILLEPELRNFPFKTKISLTQEITSLWSFNQAR-AKSOPEEYKKGFFEGALGEDD 547
DB 840 T-----LKTENLPFLTKIDTDTFFSQWSTKTKDLQKIQDLRVRYNE--LIDSS 889
QY 548 NLDFAQNTVLDK-----DYYSKKILSMKTRNKEYIHYIVQLQGDKISYEASCNL 597
DB 890 TIDFKNLYEIDQMLHMLMNMDDIAKSLFSLQVIAEKIRMT-----IPVNIINI 943
QY 598 FSKDPYSSILYOKNIEGSETAYVYVADAIEIKIDKIRIPYQISNKRNIKLTFFIGHKSE 657
DB 944 YD-----LHKNDNDLSMSIKGLAS---NPHTKINILYSNKTEHNI----- 983
QY 658 FNTDTFANLDVDSLSSEIETILNLAADISPKYIEINLLGCMFMSYSISAEITYPGKUL- 716
DB 984 FIKDLFS---FAMNELRDIIINNMKDKTP-----ENWEGRVML 1020
QY 717 -----LKIKDVSELMPSISQDSITVSAQYF---VRINEEG-KREILDHS--GKWINKE 765
DB 1021 QRYLELMKMDHLSLOSSQZANEFLISTFIYENDFLREKIEAVKKNMNSHLYFEKIKKE 1080
QY 766 ESIIDKISSKEY-----ISFNPKENKIIVSKYHLHELSTLLQETIRNNANGSDIDL 815
DB 1081 QNTWQDLSSTKEOKLIKALKEISGTEKD-----SHYDRLLDAFLK--KHNEH-----I 1128
QY 816 EKKVMLTECEINVAS-----NIDROIVEG-----RIEAKNLTSD-----SINY 854
DB 1129 HNKIQRIKDEPKESYRVAIHNDKVIKFGQTLDRLYHGVFSDINTLSRYTLHGLGNG 1188
QY 855 IKNEPKLIESIDSLYDLKHQGLDDSHFISPEDISKTEGNGPRIRFINKETGNSFIETE 914
DB 1189 VHTENLLPAPSSSLINILKE-----HYNDEISAKPLAYDY-ILNKKSESSIPVEIL 1241
QY 915 KEIFS-----EYATHISKEISNIKDTIFDNNVGNKLVKKV----- 948
DB 1242 NKLSELPPHELLTPVLGQSNPLGMGYSDNGKITEQVIVSGADGFDNPFISGLIYTYLED 1301
QY 949 --NLDAAEHVNTLNSAFFIQSLIEYNT-----TKESLSNLSVAMKVQVYQAQFS-----T 996
DB 1302 LYNHVMREGTUNSO-NURQLLENSVSCFLTEQSIKNKLLSEAEKRPQSILTEIHOHLT 1360
QY 997 GLNTITDASKWVSLVSTALDET 1018
DB 1361 GLPTIADAT--LSLLSVGLPGT 1380

RESULT 14
Q9RPH1 PRELIMINARY; PRT; 3223 AA.
AC Q9RPH1;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE EPEC factor for adherence.
GN EFA1.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E45035;
RA MEDLINE=20117988; PubMed=10652089;
RA Nicholls L., Grant T.H., Robins-Browne R.M.;
RT "Identification of a novel genetic locus that is required for in vitro
RT adhesion of a clinical isolate of enterohaemorrhagic Escherichia coli
RT to epithelial cells."
RL Mol. Microbiol. 35:275-288 (2000).
DR EMBL; AF159462; AAD49229.2; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001917; AminoTrans II.

```

DR InterPro; IPR007577; Gly_transf_sug.
DR InterPro; IPR006473; Peptidase_C58_Yt.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF04488; Gly_transf_sug; 1.
DR TIGSFAMS; TIGR01586; yopT_cys_prot; 1.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
DR PROSITE; PS00639; THIOLEPROTEASE_HIS; 1.
SQ SEQUENCE 3223 AA; 365949 MW; CAFD59F69242D07A CRC64;

Query Match 7.5%; Score 389; DB 2; Length 3223;
Best Local Similarity 19.8%; Pred. No. 9.5e-07;
Matches 254; Conservative 205; Mismatches 431; Indels 392; Gaps 55;

QY 2 NLVNAQLOQVYKVFRIQDEYVAILNALEYHNMSESS-----VVEKYLK-LKDINN 54
DB 226 NFQKISLSDELTKY---ANELIEIKRIMGEYNLLPKNSRNGKLKQKQADLLKIIME 282

QY 55 LTDNYLNTYKSGRNKALKKFKYELTMVELELN---NSLTPVERKLNLFHPIVGGQINDTA 111
DB 283 DTSVTENTPKN-----IEMAITDKREYVSTVDIEKNIHAIWVAGPPESI 329

QY 112 INVINQKQVNSDYTVKVFYDSNAFLNLTLLKKTIVESATNNTLESFRENLNDFPDYKNF 171
DB 330 SDYIKTFLTYKEFTYLLWDEKAFGAKFTSVLKQIAFDLACRTIQQTPKKNIDFINL 389

QY 172 Y---RKR-----MEIIVDKQKHFDY-- 189
DB 390 YNEIRKKNYNNPSGQOEYLNKURELYATYQKISTPLKMFNFSEFFLENMILKQDNFNICY 449

QY 190 YKQIEENPEFIID---NIIKTYLSNE---YSKOLEALNKYIEBSL-----NKIT 233
DB 450 VKGVTEINDELRLNLYLNKVIK---LSDDDIGNYQKTIINDKRVKYLILDLQKQFGENRIS 507

QY 234 ANNGNDIRNLEKFADEBLVRLNQELVERNWLAAASDIIRISMLKEDGGVYLDVLDLPGI 293
DB 508 IKQVNSLTSLSKSENN---HNYQTEMLLRWNPYPAASDLLRMVTLKEHGGIYTDITDMPEAY 564

QY 294 QPDLFVSKNPKDSITNTSHEMVKL-----EAMKYEYIPGYTSKNFMDLD--EYQVSFE 347
DB 565 SKQVIFKIMQTMGNDRFLEDLKLRRAISDGVLYR-----VNNQNIDEVNYEI----- 613

QY 348 SALSSKSDKEIEFLPDDIDIKVSPLEVKIAPANSV-----INQ 385
DB 614 ----SDADKNIIKILTEISKWEDSIFTKINTRIPDTWPIRLRVHLWPDGWNIRGLNG 669

QY 386 ALISLKDSYCDLVI---NQIKNRYKILNDLNPS-NEGTFDNTTKMIFSDKLASINED 442
DB 670 FMLSHKGEVIDAVIAGQCAVRELRRIRDNHSEIV---FKQT-----DELSLSPDPT 720

QY 443 NMFMKITYLNKVG-----FAPDVSTINLSGP----- 471
DB 721 KIGGIL-VKKYLSGSLFSKFRQDTIIPALSTLIQISGPDLIQRKMLQFRSRGVLEEF 779

QY 472 -----GVY---TCAYQ-----DLLMPKDNS 488
DB 780 NERKLSKAVIGYKVTGTGKYDNLTPESIGVNDVTPADESTWICIGKRCVDDFLPKDVS 839

QY 489 TNIHLPELRNPEFPKTKISQTEQBITSLWFSNQAR-AKSFEEYKXGYFEGALGEDD 547
DB 840 T-----IKTENLPELFTKTDITDFFSQWNTKTKQDKQKIQDLTVRYNE--LIDSS 899

QY 548 NLDPAQNTVLDK-----DYVSKILSSMKTRNKEYIHVIVQLQGDKISYEASCNL 597
DB 890 TIDFKNLYEIDQMLHMLMLENNDIADKRSLSLQVQIAEKIRMT-----IPVDNIINI 943

QY 598 FSKDPYSILYQKNIQESGSETAYYYVADAIEKIDKIRIPYQISNKNKIKLFTIGHKSE 657
DB 944 YPD-----LHKQNDNLSMSIKGFLAS---NEPTKINILYSNKTEHNI----- 983

QY 658 FNTDTFANLDVDSLSSEIETILNADKADISPKYIEINLLGNMFSYSIAEETYPGKLL- 716
DB 984 FIKDLFS---FAMNENELRDIINNSKDKTP-----ENWEGKVM 1020

QY 717 -----LKIKDRVSELMPISQDSITVSANOYE---VRINEEG-KREILDHS--GKWINK 765
DB 1021 QRYLELRKMDHLSQSSQSEANEFLEISTFIYENDFLREKIEAVKXNKHSHLYFEKIKKE 1080

QY 766 ESIKIDISSKEY-----ISFNPENKNIIVKSKYLHELSTLLQEIERNNANSSDIDL 815
DB 1081 QNTWQDLSYKEQKLQIKALKKEISNGTEKD---SHYDRLLDAFFK--KHNEI-----I 1128

QY 816 EKKVMLTECEINVAS-----NIDRQIVEG---RIBEAKNLTSF-----SINY 854
DB 1129 HNKIQRIDKFEKYSRVATHNIDKVFPGQTLDRLYHEGVFSDINTLSRYTLHLGLGITG 1188

QY 855 IKNEFKLIESISLDVLKHQGLDSDHSFISPDIDSKTENGFRIRINKETGNSIETETE 914
DB 1189 VHTPEENLLPAPSSSLNLIKKE-----HYNEDEISAKLPLAYDY-ILNKKESSIPVEIL 1241

QY 915 KEIFS-----SVATHISKEISNIKDTIFDNVNGKLVKKV----- 948
DB 1242 NKUSELPPHELLTPVLCQSVNPLMGYSSDNGKITQVIVSGADGPDNPISGLIYTYLED 1301

QY 949 --NLDAHAHYNTLNSAFFIQLSLEYNT-----TKESLNLNVAMKVQVYAQLFS-----T 996
DB 1302 LYNHIVMRREGTLNSQ-NLRQLLENSVSSCFLTEQSIKLLSEAKRFPYQSLTEIHQHLT 1360

QY 997 GLNTITDASKVELVSTALDET 1018
DB 1361 GLPTIADAT--LSLLSVGLPGT 1380

RESULT 15
Q9RM48
ID Q9RM48 PRELIMINARY; PRT; 3223 AA.
AC Q9RM48;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Lymphostatin.
GN LIPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E2348/69;
RX MEDLINE=2019489; PubMed=10722613;
RA Klapproth J.M.A., Scaletsky J.C.A., McNamara B.P., Lai L.C.,
RA Maistrom C., James S.P., Donnenberg M.S.;
RT "A large toxin from pathogenic Escherichia coli strains that inhibits
RT lymphocyte activation";
RL Infect. Immun. 68:2148-2155 (2000).
DR EMBL; AJ133705; CAB55629.1; -;
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001917; Aminotransf_II.
DR InterPro; IPR007577; Gly_transf_sug.
DR InterPro; IPR006473; Peptidase_C58_Yt.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF04488; Gly_transf_sug; 1.
DR TIGSFAMS; TIGR01586; yopT_cys_prot; 1.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
DR PROSITE; PS00639; THIOLEPROTEASE_HIS; 1.
SQ SEQUENCE 3223 AA; 365963 MW; 03B45ECDDE7938C1 CRC64;

Query Match 7.4%; Score 386; DB 2; Length 3223;
Best Local Similarity 19.7%; Pred. No. 1.2e-06;
Matches 253; Conservative 206; Mismatches 431; Indels 392; Gaps 55;

QY 2 NLVNAQLOQVYKVFRIQDEYVAILNALEYHNMSESS-----VVEKYLK-LKDINN 54
DB 226 NFQKISLSDELTKY---ANELIEIKRIMGEYNLLPKNSRNGKLKQKQADLLKIIME 282

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:34:35 ; Search time 10.5 Seconds

(without alignments)
5058:244 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1020

Perfect score: 5189

Sequence: 1 MNLVKAQLQKMYVFRIQ.....ITDASKVVELSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4101	79.0	2366	1	TOXB_CLODI
2	2791	53.8	2710	1	TOXA_CLODI
3	337.5	6.5	2867	1	RBP2_PLAVB
4	282.5	5.4	2869	1	RBP1_PLAVB
5	257.5	5.0	1162	1	EXEN_CLOBO
6	255	4.9	1875	1	MLP1_YEAST
7	251	4.8	2136	1	YCF2_MARPO
8	249	4.8	1679	1	YIO9_YEAST
9	245	4.7	1630	1	MSPI_PLAFK
10	245	4.7	1639	1	MSPI_PLAFW
11	243	4.7	1928	1	MYSL_YEAST
12	242	4.7	1005	1	RA50_METJA
13	241	4.6	1682	1	MSPI_PLAF3
14	240	4.6	1726	1	MSPI_PLAF6
15	240	4.6	1726	1	MSPI_PLAF7
16	239.5	4.6	1957	1	SPOF_SCHPO
17	239.5	4.6	3433	1	UTRO_HUMAN
18	239	4.6	1701	1	MSPI_PLAF7
19	239	4.6	3210	1	CSNF_HUMAN
20	238	4.6	1701	1	MSPI_PLAFM
21	237.5	4.6	1162	1	EXEN_CLOBU
22	237	4.6	1163	1	SPCC_CLOAB
23	236.5	4.6	1790	1	USOI_YEAST
24	236.5	4.6	2339	1	RFC1_PLAFA
25	234	4.5	1179	1	EX5B_BUCAP
26	233	4.5	1225	1	Y309_MYCGE
27	231	4.5	1169	1	SMC_METJA
28	231	4.5	1658	1	YMG7_YEAST
29	230.5	4.4	1727	1	ALM1_SCHPO
30	229.5	4.4	950	1	Y511_RICPR
31	229	4.4	1324	1	SMC4_SCHPO
32	228.5	4.4	1208	1	PCP1_SCHPO
33	228.5	4.4	1290	1	RA50_SCHPO

34 226.5 4.4 1196 1 BXCN_CLOBO
35 226 4.4 1175 1 YF19_METJA
36 224 4.3 1803 1 YUL3_YEAST
37 223 4.3 1102 1 RPOP_AGRBT
38 223 4.3 1956 1 ATX1_PLAFA
39 222 4.3 2059 1 TEGU_HSV7J
40 221 4.3 1169 1 EX5B_BORBU
41 220.5 4.2 2104 1 MY93_SCHPO
42 220 4.2 1018 1 YC14_METJA
43 219.5 4.2 1228 1 YHV5_YEAST
44 219.5 4.2 1805 1 HNW2_MYCGE
45 219.5 4.2 2663 1 CENE_HUMAN

ALIGNMENTS

RESULT 1
TOXB_CLODI
ID TOXB_CLODI STANDARD; PRT; 2366 AA.
AC P18177;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Toxin B.
GN TOXB OR TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90326540; PubMed=2374729;
RA Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
RT "Nucleotide sequence of Clostridium difficile toxin B gene."
RL Nucleic Acids Res. 18:4004-4004(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1271-2366 FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=92293124; PubMed=1603068;
RA Eichel-Streiber C., Laufenberg-Feldmann R., Sattngen S., Schulze J.,
Sauerborn M.;
RT "Comparative sequence analysis of the Clostridium difficile toxins A
and B."
RL Mol. Gen. Genet. 233:260-268(1992).
CC -!- FUNCTION: Cytotoxin.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; X531138; CAA37298.1; -;
DR EMBL; X92982; CAA63562.1; -;
DR EMBL; X60984; CAA43299.1; -;
DR PIR; A27636; A27636.
DR PIR; S10317; S10317.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF01473; CW_binding_1; 17.
DR Pfam; PF04468; Gly_transf_sug; 1.
KW Toxin.
SQ SEQUENCE 2366 AA; 269709 MW; E1024BD8B8A56ADF CRC64;

Db 120 INAEYNIKLWYDGEAFVNTLKKAIVESSTTEALQLLEBBIQNPFQDNMKFYKRMFIY 179

Qy 181 DKQKHFIDYKSOIENPEFIIDNIITKYSNYSKDLALNKIYESLNKINTANNNDI 240

Db 180 DRQKRFNYKSOINKEPTVIDIITKSHLVSEYRDETLESYRNSLRKINSRNGDI 239

Qy 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDPGLQPLPKS 300

Db 240 RANSLFTEQELNIIYQELNLRGNLAAASDIVLLAKNPGGVYLDVDPGLHSLFKT 299

Qy 301 INKPDSTINTSWEMIKLEAIMKIKYIPGYTSKNFDMLEDEVQVRSFESALSSKSKSEIF 360

Db 300 ISRPSSIGLDRWEMIKLEAIMKIKYINNTYSENFQKLDQQLKDNFKLIIESKSEIF 359

Qy 361 LPDDIKVPLEVKIAPANNVINQALISLKDYSQSDLVINQIKNRYKILNINLPSINE 420

Db 360 SKLENVSDLEIKIAPALGVINQALISKQGSYLVNLVTEQVKRYQFNLQHLNPAIS 419

Qy 421 GTDPTMTKIFSKLASINEDNMFMKITYLKVGPAPDVSTINLSGPGVYTCAYOD 480

Db 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQGVFPEARSTLSLGGPGAYASYD 479

Qy 481 LLFMKDNSTWHLPELRPEPPTKISQITQETSLWSFNOARAKSOFEBYKKGYPE 540

Db 480 PINQENITIEKTKASDLIEFKPENNLSQLTQEIINSLWSPQASAKYQFEKYVDRYTG 539

Qy 541 GALGEDNLDFQAQTVLDKDY-VSKTILSS--MKTRNKYIHYIVLQGDGKISYEASCNL 597

Db 540 GLSLEDNGVDNKNATLDKYNLNNKIPSNVBEAGSKYVHVIIQLQGDDISYEATCNL 599

Qy 598 FSKDPYSILYQKNIEGSETAYYYVAD--AEIKEDIKYRIPQIKNKNIKLTFIGHGK 655

Db 600 FSKNPKNSIIQKNV--NESAKSYFLSDGDESILELNKYRIPERLKNKEKVKVTFIGHGK 657

Qy 656 SEFTNDANLVDLSLSEIETILNLAKADISPKYIENLLGCMNMSYSISABETYPGKL 715

Db 658 DEFNTSEFARLSVDSLSEISSFLDTTKLDISPKNVEVNLGCMNMSYDFNVEETYPGKL 717

Qy 716 LLKTKDRAVSELMPSISODSITVSANOYEVINEEGKEIILDSGKWKINKEESIIKDISK 775

Db 718 LLSTMDKITSLPDKNSITIGANOYEVINEEGKEIILDSGKWKINKEESIIKDISK 777

Qy 776 EYISFNPKENKIIIVKSKYLSHELSTLQEIIRNANSDDIDLEKKVMULTECHINVASNDRQ 835

Db 778 EYIFFSDIDNKLKAKSNKIPGLASISIDITKLULLDASVSPDTKFLNLLKLNIESISIDY 837

Qy 836 IVEGRIBEAKNLSDSINYIKNEPKLTESISDSYLDLKHONGLDSDHIFISFEDISKTENG 895

Db 838 IYVEKLEPVKNIHNSIDDLIDFENLLENVSDLEYELKLNLDKYLIIISFEDISKNNST 897

Qy 896 FRIRPINKETGNSIFITEKEIFSEYATHISKEITSNIKDTIFDNVNGKLVKYNLDAHE 955

Db 898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLNLDNIQLDHTSQ 957

Qy 956 VNTLNSAFFIQLSYNTKESLNSLVAMQVYVYQALFSTGLNTITDASKVVELVSTAL 1015

Db 958 VNTLNAAFFIQLSDYSNKKVDNLDSTSVKQVLYAQALFSTGLNTIYDSQLVNLISNAV 1017

Qy 1016 DETID 1020

Db 1018 NDTIN 1022

RESULT 3

ID_RBP2_PLAVB

AC Q00759; QSN2M3; PRT; 2867 AA.

DT 01-APR-1993 (Rel. 25, Created)

DT 15-MAR-2004 (Rel. 43, last sequence update)

DT 15-MAR-2004 (Rel. 43, last annotation update)

DE Reticulocyte binding protein 2 precursor (PVRBP-2).

GN RBP-2 OR RBP2.

OS Plasmodium vivax (strain Belem).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=31273;

RN [1]

RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.

RX MEDLINE=20299192; PubMed=10838229;

RA Galinski M.R., Xu M., Barnwell J.W.;

RT "Plasmodium vivax reticulocyte binding protein-2 (PVRBP-2) shares structural features with PVRBP-1 and the Plasmodium yoelii 235 kDa zhoptry protein family.";

RL Mol. Biochem. Parasitol. 108:257-262(2000).

RN [2]

RP SEQUENCE OF 1189-2439 FROM N.A.

RX MEDLINE=92315338; PubMed=1617731;

RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;

RT "A reticulocyte-binding protein complex of Plasmodium vivax merozoites.";

RL Cell 69:1213-1226(1992).

CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to human reticulocyte cells.

CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).

CC EMBL: AF184623; AAF76525.1; -

DR HSP; P03069; IGM; Transmembrane; Repeat.

KW Malaria; Receptor; Signal; POTENTIAL.

FT SIGNAL 1 21

FT CHAIN 22 2867 RETICULOCYTE BINDING PROTEIN 2.

FT DOMAIN 22 2867 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 2806 2826 POTENTIAL.

FT DOMAIN 2827 2867 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 44 133 ASN-RICH.

FT DOMAIN 560 758 LYS-RICH.

FT DOMAIN 1112 1285 LYS-RICH.

FT DOMAIN 2758 2785 7 X 4 AA TANDEM REPEATS OF H-D-D-T.

FT REPEAT 2758 2761 1.

FT REPEAT 2762 2765 2.

FT REPEAT 2766 2769 3.

FT REPEAT 2770 2773 4.

FT REPEAT 2774 2777 5.

FT REPEAT 2778 2781 6.

FT REPEAT 2782 2785 7.

SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFFD3 CRC64;

Query Match 6.5%; Score 337.5; DB 1; Length 2867;

Best Local Similarity 19.8%; Pred No. 1e-05;

Matches 250; Conservative 226; Mismatches 454; Indels 335; Gaps 57;

QY 2 NLV-NKAQKQKVVYKFR-----IQEDYVAILNALALEEYHNMS----- 40

Db 977 NIINKKADILKLDQHTQIEKHFTIENEEMSPLLSVIKKEKNRVESDWSELIKQLNTK 1036

QY 41 --SVVEKYLKLD-----INNLTNYLNTYKSGNKAL 72

Db 1037 INAILLEYNKSQDRFNGDDTNLEELDDFKKQDQAQOEIKKLTNY--NVLNDGINVII 1094

QY 73 KKFKEYLTWVLELKNNSLTPEKXNLHFIWIGQINDTA---INVINQWK-----DVNS 123

Db 1095 KEQHE-----KVIILSENHITKDK-----KINEKIQQNVNLSNEMKTKLGLLKINE 1141

QY 124 DVTVKVFDSDNAFLINTLKTIVE-----SATNNTLESFRE--NLNDPEFDVN 169

Db 1142 D--IKNSRD-----TKSKIQEKKVQVTFGSIDYANKKIDAKKEHDVNDKDFDKE 1193

QY 170 KF-----YRKMEIIVDKQKHF--DYKQSIQENPEFIDNIITKYSNYSKDLAL 221

Db 1194 KYKDTSFDEKKKSIKAYEKGKNTLKELEKMDDEKNIEKEVEAQIQYKRFIDHVNLM 1253

```
QY 222 NKVIESLNKITANNND:RNLEKFADELDVRLYNQEL-----VE 261
Db 1254 NDEVKS--KIV-----MEK-----IELYKEIDEIKQKNEYKQDGNFYFTE 1296
QY 262 RWNLAAS-----DILRISMLKE-DGVVYLDVILPGIQPDLFKSPKDPISITNSWEM- 314
Db 1297 QYNSATQSKAKIEQFINIATTKGTSDTSODINELLESKEEVHKLQLVKQESNSMEBR 1356
QY 315 -----IKLEIMKYKEIPIGYTSKN-----FMDLDEVQVSFESALSSKSD 355
Db 1357 KQILSKMDLLLNNSSETIAKESNTQNALGFRENAKTKLNKTDELLQKVAAMTEAAH 1416
QY 356 KSEIFLPDDIKVSPLEVKIARANSVINOQ-----LISLKD--SYCSDLVINQIKRY 407
Db 1417 KNNIDIALEDAQIDTEVSKIEQINREIMNKDEIKSYLSEIKYKDKCTTISNSKRGKD 1476
QY 408 KI-----LNDNLNPSINEGTDF-----NTWKIFSDKLASIS 439
Db 1477 KIEFLKFKPNBESNKNVINEINENIRNSQYIKDIEDAKQASTKVELFKHETIIS 1536
QY 440 N--EDNMFMKITYNLYKVPADVRSTINLSGPGVYT--GAYQDLLMFKONSTNIHLE 495
Db 1537 NIFKESILGVETKSKINKAEDIMKEFERHENSEIQTQVKGQFQENKLNKNEPHYDNE 1596
QY 496 PELRNFEPTKISQTEIITSLMSFNQARAKSQFEYKGYFEGALGE---DNLDPA 552
Db 1597 DELNNDKSTNAKV--LIETNLSV-----KHNLSITNIKQ-----GEKIYSKAKIM 1643
QY 553 QNTVLDKDYVSKKILSSMKTRNKYIHYIVQLQGDKISYEASCNLFPSKDPYSSI--LYQKN 611
Db 1644 QKIKATSENTAEKTEKVKDDQSNVNYL-----NQITERNLIVTEKRLNGIDSTITN 1698
QY 612 IEQS--ETAYYYVADAE--IKEIDKVRIPQISNKRNIKLTIFHGKSEFNT-----DT 662
Db 1699 IEGALKESKNGVEIGLEKLEIGKRLKVDITKKSINST-VGNFSSLFNNFDLNQYDF 1757
QY 663 FANLD-----VPSLSSEIETILNLAADI-----SPKYI-----EINLLG-- 697
Db 1758 NKUINDYENKMGIEYNEFGSLNKISENLARNASENTSDYNSAKTLRLAEQKKNLANK 1817
QY 698 --CNMFSYSISABETYPGKLLIKIDRVSELMPSISQDSITVSANQYVRINEEKRIL 755
Db 1818 EEANKYLRDYKVESF--RFIFNMKESLDKINEMIKKEOLTYNBEGHGNVQLVENEKILV 1875
QY 756 DHSGKWKNEESTIKDISKEVYSFNPKNKILIVSKYLHELSTLLOERNNANSDDIDL 815
Db 1876 DE-----NLSLDLKQATGN-----EEIQKITHSLKKNKAKTILGHVDTSAKVGIKI 1924
QY 816 EKKWMLTE-----CEINVASNIDRQIVEGRIEBAKNLTSDSINYIKNEFKLIBESID- 867
Db 1925 TPELALTELLGDAKLTAQELKFBESKNVVLATENMSKT-----NELDVHKNIODAYKV 1979
QY 868 SLYDLKHQGLDSDHIFSEDISK-TENG-----FRIRFINKETGNSFIETEKETFSYA 922
Db 1980 ALEILAHSDIDPKQ---KDSKLEMGNQIVLVKVLINQYKNKISSIKSKEEAVSVKI 2035
QY 923 THISK--EISNI--KDTIFDNVNGKLVKKNLDAHEVNTLNSAFFIQ-----SLI 969
Db 2036 GNVSKHSELSKITSKSDKYDNT-----IALEKQTELQNLNSFTQEKNTNSDSKLE 2088
QY 970 EYNTKESISNLSVAMK-----VOVYAQLFSTGLNTITDASKVBLVSTA 1014
Db 2089 KIKTDFESLKQALKTEGEVNALKASSDHNHEVQSKSEPNPALSIEIEKEETDIDSLNTA 2148
QY 1015 LDETI 1019
Db 2149 LDELL 2153
```

RESULT 4
RBPI_PLAYB
ID_RBPI_PLAYB

```
AC Q00798;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Reticulocyte binding protein 1 precursor.  
GN RBPI.  
OS Plasmodium vivax (strain Belém).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI_TaxID=31273;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92315338; PubMed=1617731;  
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
RT "A reticulocyte-binding protein complex of Plasmodium vivax  
merozoites.";   
RL Cell 69:1213-1226(1992).  
CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to  
human reticulocyte cells.  
CC -!- SUBUNIT: Homodimer (Potential).  
CC -!- SUBCELLULAR LOCATION: Membrane-bound.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
DE EMBL: M88097; AAA29743.1; -;  
KW Malaria; Receptor; Signal; Transmembrane.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 2869 RETICULOCTE BINDING PROTEIN 1.  
FT DOMAIN 18 2807 EXTRACELLULAR.  
FT TRANSMEM 2808 2826 POTENTIAL.  
FT DOMAIN 2827 2869 CYTOPLASMIC.  
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).  
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205BECFF CRC64;  
  
Query Match 5.4%; Score 282.5; DB 1; Length 2869;  
Best Local Similarity 20.0%; Pred. No. 0.0014;  
Matches 242; Conservative 212; Mismatches 410; Indels 349; Gaps 61;  
  
QY 6 KAQLQKVVYKFRIOBDEY-----VAIINALEEYHNMSESVVVKYKLLK-DINNL- 55  
Db 946 KALKEKIVSDSLDKIDQYETEFKEKTSVAVENTVSTIQSLSKA--IDSLKRLNGSINNCK 1003  
QY 56 ---TD-NYLYNTYKSGRNKALK-----KFKBYLNEVLELKNLSLTPVKQLHFIWIG 104  
Db 1004 KYNTDIDLLRSKIKTLREEVQKEMPKRGDKCGENTTALLKSLDRKMGKINEXLN---D 1059  
QY 105 GQIN--DT-----AINVINQWQVNV----- 122  
Db 1060 GRNLSTDKKEDLLKFYSESKSIHLKSKQKQGPDPANRIDEWEDIKRDVDELNVYQVI 1119  
QY 123 SDYTVKVFYDSNAFLNTLKKTI-----VESATNTNTLESFRE-----NLNDPEFYDNKF 171  
Db 1120 SENKVTLFKQNSVTYIEAMHSHINTVAHGITSNKLILSKVEDEKLNLEQVEDYKVV 1179  
QY 172 ----YKRMEEII---YDKQKHFDIYKYSQIEENPEFIIDNIKIYLSN-----EYSKLEA 220  
Db 1180 KNPENEQLEAIRGSMKLEKINKHVSEMTQ-----LESTANTLAKNAGKENBHDLEE 1234  
QY 221 LNKY-----IEESLNKI-----TANNGNDRNLEKFADELDVRLYNQELVERNLAA 267  
Db 1235 LNKTKGQMRDIYEKLLKIABELKEGTVNELKANEKANKVEPEPEPNIIGHVLE----- 1288  
QY 268 ASDILRISMLKEDGVVYLDVILPGIQPDLFKSPKDPISITNSWEMIKLEAIMKYKEYI 327  
Db 1289 ----RITVEKQKAGKV--VEEMNSUKTIEKLIQE---TSDDSQNELVTTSTIKHLNA 1338  
QY 328 PGVTSKNFMDLDBEVQVSFESALSSKSDKSEIFLPLDDIKVSPLEVKI----AFANNSVI 383
```

```

Db 1339 KGY -----EDVIKNEEDSIQLR-EKAKSLETLDEMKGVLQVQNMNLQSAIQGNAGI 1389
Qy 384 NQALISLKDSVCSLVLNQIKRKYILNDNLNSINEGTDFTMTKIPSDKLA-----S 437
Db 1390 SKEL-----NEUGVIELLISYNSILEYVKKNSSESVRFSQLANGFTKA 1436
Qy 438 ISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGVVTGAYQDLLMFKDNSTNIHLEPE 497
Db 1437 EGEKNASARLAEBKLEQIVKDL-----DYSB-----IDDK 1469
Qy 498 LRNFEPKTKISQTEQETISLMSFNQARAKSOFEEYKGYFGALGEDNDLDFANQ--- 554
Db 1470 VKIEGIGIKRELKWKESALT-FWESEB-----KFKQKCSHMEVAKEGKKIEYLVKNGD 1523
Qy 555 -----TVLDK-----DYVSKILSSMKTRNKEVHIYVQLQGDKISYEASC----- 595
Db 1524 GGRKANITDSQMEVGNVTSK-----AEHAFHTVEAQVDKT-KAPCESIVAYVTK 1571
Qy 596 --NLFSKDPYSSILYQKNIEGSETAYYYYVADAEIKEDKRYIPYQIS-NKRN-----K 647
Db 1572 MDNLFNSLMKEVKVCKEKNDAEKY-----SAKIKPYDG-RIKARVSENERKISLKKK 1626
Qy 648 LTFIGHGKSEFN-TDTPANLDVDSLSSEIETILNLAKADISPKYIIBNLGCMNFSYSIS 706
Db 1627 AKVEKSSQLNDVSTSLQIDNCRQQLDSVLS-----NLGRVKQNALQYFDS 1675
Qy 707 ABETYPGKLLKIDRVSELMPSISQDSITVSANQVEVRINEREGKEIILDHSGKWINKEE 766
Db 1676 ADKSMKSVL-----PISELGAESKLDKVAAKESYEKDL-----ETQNMESRINVEE 1723
Qy 767 SIIXDISKEYISFNPKENKIIVSKYHLHELSTILLQEIIRNANSDDIDLEKKVMECEI 826
Db 1724 GSLTDIDKK-----ITDIENDLLMKKQYEE--GLLOKIKENADKRSNFE-----LVGSEI 1773
Qy 827 NVASNIDRQI-VEGRIEEAKNLTSDSINY-----IKNEFKLTESDSLYDLKHONGLD 879
Db 1774 NALLDPSTSIPIKULKE-YDMTGDLKNYGVKNEIHGEF-----TKSYNDIETHLSNAYD 1828
Qy 880 DSHFISFE-----DISKTENG-----FRIRFINKETGNSIFITEKEIFSEYA- 922
Db 1829 YS--VTFEKAQSLRELAKEEHLRRREBEAIFILNDIKKVESLKLKEMKVKVSAEYEG 1886
Qy 923 -----THISKEISNIKDTIFDNVNG-----KLVKVY-----NLDAEHEVN- 957
Db 1887 MKRDHTSVSLVQDMK-TIVDELKTLNDISECSSLNNVSVIVKVKESGHADYRRDANS 1945
Qy 958 -----TINSAFF-----IQSLIEYNTKESLNSLVAMKVQVYQI-FSTGLNTIITDASK 1006
Db 1946 MYESMVTLANVFLSDEAKISGMEFNA--EMKGNFKTDLELEIF-----SVISNSNE 1995
Qy 1007 VVELVSTALDETI 1019
Db 1996 LKKIEQDSNDVI 2008

```

RESULT 5

```

BXEN CLOBO
ID _BXEN CLOBO STANDARD; PRT; 1162 AA.
AC P46082;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type E, nontoxic component.
GN ENT-120.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Maslike;
RX MEDLINE=93195515; PubMed=8450310;

```

```

RA Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT "The complete nucleotide sequence of the gene encoding the nontoxic
component of Clostridium botulinum type E progenitor toxin.";
RL J. Gen. Microbiol. 139:79-86(1993).
CC !- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
TOXICITY.
CC
CC This SWISS-PROT entry is copy-right. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.ebi.ac.uk/ebisib/ or
send an email to license@ebi.ac.uk).
CC
CC EMBL; D12697; BAA02194.1;
CC InterPro; IPR008985; ConA_like_leg_gl.
CC InterPro; IPR000395; Peptidase M27.
CC Pfam; PF01742; Peptidase M27; 1.
CC PRINTS; PR00760; BONTOKILYSIN.
CC ProDom; PD001963; Bontokilysin; 1.
CC Neurotoxin.
CC
CC SEQUENCE 1162 AA; 136856 MW; 96468EDDDAB0F39D CRC64;
Query Match 5.0%; Score 257.5; DB 1; Length 1162;
Best Local Similarity 19.8%; Pred. No. 0.0042;
Matches 229; Conservative 171; Mismatches 408; Indels 349; Gaps 52;
Qy 34 YHNMSESSVWRKYKLKD-----INNLTNDYNTYKSGRKNALK-- 73
Db 109 YENNTEDYRQNTYLLSSKNEHYVTANLVIPOGSGNIKNVYIYKYEAGSGMTLBIW 168
Qy 74 -----KFKYLTWEVLEL-----KNSLTPVEKNL 98
Db 169 POPFLTHKYDFYVDPALCLKIKLSLYLYGKPNLNLPIYRLRNEFNSLEYSELNM 228
Qy 99 HEIWIGGQINDTAINYNQW-----KDVNSDYTVKVYDYSNAILNTLTKTI 145
Db 229 IDFLISGGIDYKLLNTPNYFWIDKYFIDTSKPFKYNDYFIKI--KNNYIANSIKLYL 286
Qy 146 VESATNTLTSEFRENLN--DPEFD-----YKPYKRMELI-----YD----- 181
Db 287 BQKFKINVKDWEVLSVFSKEFQIMPERYNNALNHYRKEFYVIDYFKYNNYNGPKNG 346
Qy 182 --KQKHFDIYKQTEENPEFIIDNI-----IKTYLSNEYSKOLEALNK 223
Db 347 QIKTKLPUSKYNKEILNKPELIVNLNQNTVLMKSNYIGDLKGTVDNFVSNVLIPTNL 406
Qy 224 YIEBSLKNITANNGN--DIRNLEKPAEDL--VRLYNQELVERWNLAASDILRISMLKE 279
Db 407 NYEHSINYFLDNVNIIEIKIPIINDEDIYPRKNADTFIPVYNTYKAKEINTTLPV 466
Qy 280 DGGVLDVLDILP-----GIQDLPKYSINKPDSITNTSWEMIKLEAIMKYKEYI----PGYT 331
Db 467 N---YLOQMIDSDNDINLSSDFLKVISKGLSVSF-----LNNMTDYLEIFKYDKPIDT 518
Qy 332 SKNFDMLDEEVQRFESALSSKSKSEIFLPLDDIKVSP-----LEVKIAFANNSVINQ--- 385
Db 519 DKYKWKALKAFRNVSLDITETQISNOF---GDKIIPWIGRALNLTNTNSFVEEPKN 575
Qy 386 ----ALISLKDSYCSDLVINQIKRYKILNDNLNPSINEGTDFTMTKIPSDKLAISNE 441
Db 576 LGPISLINKKE---NITIPKIK-----IDEIFSSM-----LNFSPKDLSENLFNIYCK 620
Qy 442 DNMMFMKITYLVKVGFPADVRSINLSGPGVVTGAYQDLLMFKDNSTNIH-----LLE 495
Db 621 NNFYLKIIYYNFDQWMTQ-----YYSQVFDLICMSKSVLAQEKLIKLIQ 667
Qy 496 PELR-----NFEFPKTK-----ISQTEQETISLWSFNQARAKSQREE--YK 535
Db 668 KQLRYLMENSNISSNLILINTLTNTLURDISNQSQTAINNIDKFFNNAMCVFENNITYP 727
Qy 536 KGYFEGALGEDNDLDFANQTVLDKDYVSKK-ILSSMKTRNKEYIHYIVQ-----LQGDKI 589

```

```

Db 728 K-----FTSMEQCKININKSTKEFLKCTININETEKSHLMQNSFSLNDFDL 776
Qy 590 SVEASCNLFKDPYSSILYQKNGESATAYYY-----VADAEIK-----EIDK-VRIP 637
Db 777 DIQNMKNLFNL--YTELLIKSQTSPYELSLYAFQEQDNNVIGDTSGKNTLVVEYFKDGLV 834
Qy 638 YOISN-----KXNIKLT--FIGHKSSEFTDTFA-----NLDVDSLSSEI----- 675
Db 835 YGINNATHGANCQNIKFTNDYFENG-----TNNFSIYFWLRNLKQNTIKSLKIGSKED 890
Qy 676 -----ETILNLAKADISPKYIEINLLGCMFMSYSISASAEETYPGKLLIKKIDORVS 724
Db 891 NCGWEIYFENDGLVFNIDSNGENKQNYLSNKNKSWHYIVISINRLKQDOLLIFI----- 945
Qy 725 ELMPISODSITVSANQVEVRINECKREIILDSKWKINKEESIKDISKEYISFNPK 784
Db 946 -----DNILVA-----NEDIK-EIL-----NIYSDIISLSDN 973
Qy 785 NKIIYKYLHLSLTLQEIIRNA-----NSSDIDLEKKVMLTECEINVASNIDR 834
Db 974 NVN-----YIEGSLVNLKTIINSNEILTYFSDLNNSYIRNFDIELQVNRTYELFNVP 1028
Qy 835 QVEGRIBEAKNL-----TSDSINYIKNEFKLIESISDSLYDLKHQ-----NGLDSDSHFIS 885
Db 1029 EIAINKIEQNNIYLSINNENNLNFKPLKLLNTNPNKQYVQKWDEVIFSVLDGTE--K 1086
Qy 886 FEDISKTENGFFIRINKETGNSIFIEFTEKIFSEYATHISKEISNIKDTIFDNVNGKLV 945
Db 1087 YLDISTNN--RIQLVDNKNNAQIFI-INNDIF-----ISNCULTYNNVNVYLS 1133
Qy 946 KK-----VNLDAAEV 956
Db 1134 IRNQDYNWVICDLNHI 1150

```

RESULT 6

```

MULTI_YEAST
ID MLPI_YEAST STANDARD; PRT; 1875 AA.
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLPI.
GN MLPI OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8483450;
RX STRAIN=S288c;
RX MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Borstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UB12 and MPL1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: Myosin-like protein that is probably involved in DNA
CC repair.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: Ref.2 misquotes the gene name as "MPL1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L01992; AAA34783.1; -
DR EMBL; X73541; CAAS1948.1; -
DR EMBL; Z28320; CAA82174.1; -
DR PIR; S38173; S38173.
DR GenOnline; 140074; -
DR SGD; S0001803; MLPI.
DR GO; GO:0005635; C:nuclear membrane; IDA.
DR GO; GO:0005654; C:nucleoplasm; IDA.
DR GO; GO:0006066; P:protein-nucleus import; IDA.
KW Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 4.9%; Score 255; DB 1; Length 1875;
Best Local Similarity 17.4%; Pred. No. 0.0094;
Matches 234; Conservative 241; Mismatches 432; Indels 436; Gaps 55;

Qy 14 YVKFRIQDEYVAIL-----NALEYVHNMSSESVVEKYLKLDINN 55
Db 225 YLSYRKTKVLDIRNELNRLNDFQWERTNNDVLKQNNLSKSLQEKLEIK---GL 281
Qy 56 TONYLNTYKSGRNKALKKFKYLTMEVLEKNSLTPVEKNLHFHWIGGQINDAINY 115
Db 282 SDSLNSEKQEPSNEMSLQ-----RLVDLLESQNLNAKVEL----- 317
Qy 116 NOWKQVNSDYTKVFD-----SNAFLINTL-----KKTIVESATNVLBSF 157
Db 318 NSIRELN--TAKVIADDSKQTPENEDLLKELQTLKEKLAQCEKELSLITDEAD 374
Qy 158 RENLNDPEFDYKFKRMEIYDKQHFIDYKQISENPEFIID-----NIKTY--- 209
Db 375 NENLSAKSSDFILKKQL-----IKERTKHELQWQIE--TFIVELEKHVFIINSFK 427
Qy 210 ---LSNEYS-----KOLEALNKVIEESLNK----- 231
Db 428 TDMLENELNNAALLLEHTSNEKNKAKVELNAKNQKLVCEENDLQTLTKORLDCRQI 487
Qy 232 -ITANGND-----IRNL-----EKPADELVLRYNQBELVRNWLAAASDILR 273
Db 488 LIITNSVNSDKGFLRKEEIQFIONIMQEDDSTITESDSOKVYTERLVEFKN----- 538
Qy 274 ISMLKEDGGVLDVILPGIQDPLFKSINKPSIINTSWEMIKLEAIMKYKYEIPLYTSK 333
Db 539 IIQLEKNAELLKV-----VRNLADKLESKEKSKSLQKIESETVNEAKEAIIITLKE 592
Qy 334 NFDMLD--BEVQSFESALSSKSKSEIFLPLD-----DIKVSPLSVKI----- 375
Db 593 KMDLESRIEELQKELEELTKVPNEDASVSNVTIKQLTETKDDLESQVODLQTRISQ 652
Qy 376 -AFANNSVINQALISLSDYSCDVLNQIK-----NRYKILNDNL-----NPSIN 419
Db 653 ESTENSLNKEIQDLYDSK-SDISIKLGKSSRLAEERFKLSLNTLDTLKAENDQLR 711
Qy 420 EGTDP--NTTMKIFS-----DKLASIGNEDNMFMWIKITNLYKVGFPADVRS 464
Db 712 KRFDYQLNTILKQDSKTHETLNEYVSCSKSLSVETE--LLNLKEQKLRVHLEKNLQ 768
Qy 465 TINLSPGVVTGAYQDLMFKONSTNIHLPELRNFPFKTKISOLT-----EQEITSIW 520
Db 769 ELNKLSP-----EKSLRIMVTQLQKEREDLEETRSQCKKIDEL- 812
Qy 521 SFNQARAKSQFEBYKGYFEGALGEDD--NLDPAQNTV--LDKDYVSKKLSLSSMKTRNKE 576
Db 813 --EDALSELKQTSQKDHKIKLEEDNNSNIEWQNKIEALKADYESVITSVDKQTDLE 870

```


QY 779 SFNPKENKIIVKSK-----YLSHLSLLQEIIRNNSSDIDLEK 817
Db 934 SFIEKKNLFIKQSUNVLLVKNYSKFFDNIFHLQKKNIEIILNNQNYFEKSLK 993
QY 818 KVMLTECEINVA-SNIDROIVEGRIEEAKNLTSDSINTVKNKFLIESDSLSY-----D 871
Db 994 KYVLKNLNNYSKFSYKIF---IFQLLNILN-----KNNYKTFQWISLFIYKXNLN 1044
QY 872 LKHQGLDSDHSPISPEDISKTENGFR-IRPINK-----ETGNSIFETEKEIFSEYATHIS 926
Db 1045 YKIQNKIEKNYCNKYNISYKKKIKTVNFFKKNLFTNNSWFFTL-----WWEYNTVIL 1101
QY 927 KEISNIKDTIFDNV-----GLVKKVNLDAAEVNTLSARFIQSILIENTIKESLS 979
Db 1102 LQI--IQETFFQITDVLEVFYKKKIKIEK-----NLKFFLKS-----KKISLK 1141
QY 980 NLS 982
Db 1142 TLS 1144

RESULT 8
YIO9 YEAST STANDARD; PRT; 1679 AA.
AC F40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 195.1 kDa protein in DNA43-UBI1 intergenic region.
GN Y1149C.

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313266; PubMed=9169870;
RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
RA Moulle S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA Skelton J., Smith V., Walsh S., Whitehead S., Barrall B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX";
RL Nature 387:84-87(1997).

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; Z38059; CAA86129.1; -.
CC PIR; S48385; S48385.
CC GernOnline; 139684; -.
CC SGD; S000141; MLP2.
CC GO; GO:0005635; C:nuclear membrane; IDA.
CC GO; GO:0005654; C:nucleoplasm; IDA.
CC GO; GO:0006006; P:protein-nucleus import; IDA.
CC KW Hypothetical protein.
SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202DBF CRC64;

Query Match 4.8%; Score 249; DB 1; Length 1679;
Best Local Similarity 19.7%; Pred. No. 0.014;
Matches 226; Conservative 193; Mismatches 373; Indels 356; Gaps 61;

QY 3 LVNAQLQKVMYKFRIOQDEYVAINLAEEVH-----NMSSESVVEKYLKLDINNLT-D 57
Db 469 LNTSAIQE---TASPLSQDELISLRKILESNNVNNDSQAIITRLVFEFNNVNELOEK 525

QY 58 --NYLNTYK-----KSGRNKALKKPKBYLTME-----VLELKNNSLTPVERNLHFIW 102
Db 526 NVZLLNCIRLADKLENYEGKQDKTLQKVENOTIKEADAIIELEN----- 571
QY 103 IGOQIN---DTAINYNQWKNVSNVYKVFYDSNAFLINTLTKTIVESATNNTLESFRE 159
Db 572 ---INAKMETRINILLRERD-----SYKLLASTEENKANTNSVTSMEAREKKIRELEA 622
QY 160 NLNDPFDYKFKRMEIYYDKQKHFIDYKSOEENPEFIDNIIKTYLSNEYSKDLE 219
Db 623 ELSSTKVE-----NSAIQNLKELLTYKSSQCKK-----KTTL----- 656
QY 220 ALNKYTELSNKLITANNNGDIRLEKFADEDLVLYNQBELVERWNLAASDILRISLKE 279
Db 657 -----EDFENFKGLAKEK-----ERM-LEEAIDHLKAELEKQ 687
QY 280 DGGV--YLDVDILPGIQDPLFKSINKPDSITNTSWEMIKLEA---IMKYKEYIPGYTK 333
Db 688 KSWVPYSIHVE-----KERASTELSQSRIKISLEYSKLEKK-----ETA 728
QY 334 NFDMLDEEVORSPESALSXSXSSEIFLPLDDIKVSPLEVKIAPANNVINQALISLKOS 393
Db 729 SFPTKESLTRDFEQCC---KEKELQMLKSEIISHNENKMDFSKSGQYKA----- 778
QY 394 YCSDLVINQIKRYKILNDNLNPSINE-----GTFDNTTMKIFSDKLAS 437
Db 779 -----KIKELENLRLSDLOSKIQEISIRCKDSQLKWAQNTIDDTMKWKS-LLTE 832
QY 438 ISNEDNMFMIM--KITNLYKVGPAFVRSTINLSGPGVVTGAYQDILLMFKNSTNHLE 495
Db 833 LSNKETTIEKLSSEIENLDK-----ELRKT-----KFKQ--YKFLDQNSDASTLE 874
QY 496 PEIRNEFPFKTKI-----SQL-TEOEITSLWSFNQARAKSQPEEYKGY----- 538
Db 875 PTLRK-ELQIQVLKQDANSIQAYEEIITSSNENALIELKNELAKTKENYDAKIELEKKE 933
QY 539 -----FEGALGE-----DDNLDFAQNT-----VLDK----- 559
Db 934 KWAREEDLSRLGELGEIRALQPKLKEGALHFYQQSEKLRNEVERIQRMKIEKMTIV 993
QY 560 DYVSKTILS---SMKTRNKYIHVIQLOQDKISYEASCNLPSKDPYSSILYOKNIEGSE 616
Db 994 QLCCKKMSQYQTMKENKDLSELVIRLEKDAADCAELTKTKSSLYSA---QDLDKHE 1050
QY 617 TAYYYYVADAE-----IKEIDKYRIPYQI-----SNKRNIKLFTFHGKSEF 658
Db 1051 RKWMEERADYERELISNIEQTESLRVENSVLIEKVDDTAANNQDKHKLIVSL----- 1103
QY 659 NTDTFANLDVDSLSSEIETILNLAADISPKYIE-----INLLGNMFSYSISAE 709
Db 1104 ---FSLRHE--RNSLETMLTKCKELA--FVKQKNSLEKTINDL---QRTQTLSEKE 1152
QY 710 TYPCKLLLLKIDRVSELMPSISQDSITVSAN---QYEVRAINBEGKEILDHSGKWINKEE 766
Db 1153 YQCSAVII---DBFKDITKEVTQVNIILKENNAILQKSLKNVTEKNEIY---KQNDKQ 1205
QY 767 SII-----KD-ISSKEYISFNPKENKIIVKSKYLHELSTLLQEI---RNNANSSDID-LEK 817
Db 1206 EETSLRQDLIQTKEQVSIN--SNKILVYSEMEQCKQRYQDLSQOQKQAKQKIEKLN 1263
QY 818 KVMLTECEINVASNIDRQIVEGRIEEAKNLTSDSINYIK-----NEFKLIESISDSL- 869
Db 1264 EISDLKGLSSAENANADL--ENKFNLKQKAHEKLDASKKQQAALTNELNKAIKDKLE 1322
QY 870 YDLKHQNG---LD---DSHFISFEDISK--TENGFR-----IRFINKE-----TNSIF 910
Db 1323 QDLHFENAKVIDLTKLKAHELOSEDSVRDHEKDTYRTLMEELESUKKELQIFKTANS-- 1380
QY 911 IETEKEYSFVATHISKEISNIKDTIFDNVNGKLVKVNLDAAAEVNTLSAPFIOSLIE 970
Db 1381 ---SSDAFEKLVNMEKE-----KORIIDERTKEFKEKQ-----ETLKS---TSSEAE 1424
QY 971 YNTTKESL 978

QY 937 -----FDNV-NGKLVKVNLDAAHEV-----NTLNSAPFIOSLIEYNTTKESL-SNLS 982
DB 1309 RPKNGKFNKVLSDLIIPKYLDTSSNYVVKDPYFLNKKEDKFLSSYNIKSIDTDIN 1368
QY 983 VAMKVQVYAQLFSTGLNTITDASK 1006
DB 1369 FANDVLGYKILSEKYSKSDLSIK 1392
RESULT 10
MSPI PLAFW
ID MSPI PLAFW STANDARD; PRT; 1639 AA.
AC P04933;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, last sequence update)
DT 10-OCT-2003 (Rel. 42, last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMSEA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI TaxID=5948;
RN [1] _
RP SEQUENCE FROM N.A.
RX MEDLINE=86014355; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA Nicholas S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites.";
RL Nature 317:270-273 (1985).
RN [2]
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -! SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -! PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC kDa and 19 kDa antigens which are the major surface antigens of
CC merozoites. The maturation take place during schizont.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X02919; CAA26676.1; -
DR PIR; A24594; A24594.
DR PIR; S05603; S05603.
DR PDB; 1CBJ; 28-MAY-99.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor; 3D-structure.
FT SIGNAL 1 19 POTENTIAL
FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 764 764 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 768 768 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. .) (POTENTIAL).

SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;
Query Match 4.7%; Score 245; DB 1; Length 1639;
Best Local Similarity 21.1%; Pred. No. 0.019;
Matches 258; Conservative 211; Mismatches 425; Indels 330; Gaps 68;
QY 16 KFRIOEDY-VAILN-ALIEYHNMSESSVVEKYL-----KLKDINLTNYINTYK----- 64
DB 275 KKKLYQAQYDLSIYNKQLEEAHL--ISVLEKRIDTLAKNENIKELLDK-INIKNPPPA 331
QY 65 KSG-----RNKAL-----KKFKEY-----LTMEVLEL-----KN----- 88
DB 332 NSGNTPTLLDKNKXIEBEHEKEIKIAKTIKFNIDSLFTDFLELEYIRREKKNIDISAK 391
QY 89 -----NSLT-PVEKNLHFIWIGGOINDTAINYNQWQDVNS--DYT---VKV 129
DB 392 VTKKSTSTNEYPNGVTIPLSYN-----DINN-ALNELNSFODLNPDPYTKPSKN 442
QY 130 FYDSN--AFLINTLKTTI-----VESATNNTLESFRENLANDPDPYKFKRMEIYD 181
DB 443 IYTDNERKFFINEIKIEKKIES--DKSYEDRSKSLN---DITKEYEKLNEIYD 497
QY 182 -KQKHFDY--YKSOIENPEFIIDNII--KTYLSNEYSK-DLEALN---XYIEE-SLNK 231
DB 498 SKFNNNIDLTNFERKWKRYSYKVEKLTHNTFASYENSKHLEKLTALKYTMEDYSLRN 557
QY 232 ITANN-----GNDIRNLEKFADEDLVRVYNQELVERWN-----LAAAADILRI 274
DB 558 IIVKELKYKKNLISKIENIEITLVENIKDBEQLEFKKIKDKENKPEKILEVSDIVKV 617
QY 275 SM-----LKEDGGVLDVDILPGIO-PDLFKSINKPDSITNTSWEMIKLEAMK 322
DB 618 QVQKVLMMKIDELKKTQILKNVELKHNHVPNSYKQENKQEPY-----YLIVLKEIDK 673
QY 323 YKEYIPGYTSKNFDMLEDE--VQSFSAISSKSDKSEI-----FLPLDD 365
DB 674 LKVFMPKVS-----LINEEKNIKTEGSDNSEPSTEGEITQATTKPGQQAAGSALEGS 729
QY 366 IKVS-----PLEVKI-----AFANKSVINQALISKDSVCSDLVINQIKN 405
DB 730 VQAQAEQKQAQPPVPVPVPEAKAQVTPPAPVNNKNTENVSKLDYLEKYLEFNTSYICH 789
QY 406 RYKILNDNLNPSINEGTDFTNTMTKIFSDKLASINEDNMFMKITNYLKVGFAPDVRST 465
DB 790 KYLVS-----HSTNNE--KILQYKTKEEESKSCDPLDLFNQNNIPWYS--MPS 842
QY 466 INLSGPGVYTGAYQ-----DLLMFKDNSTNIHLPELANFPFKTKIS-----QLTQ 514
DB 843 LNNLSQLFWEIYEKEMVCNLYKLDNDKIKNLEBAKKVSTSVKTLSSSSMQPLSLTPQ 902
QY 515 EITLSMFSNQARAKGOFEBYKGYFGALGEDDNDLDFQNTVLDK--DYVSKILSSMKT 572
DB 903 DKPEV-SANDTSHSTNANSLKLPENILSLGKNKIYQELICQKSSSENFYKILKDSOT 961
QY 573 RNKEYHYIVOLQGDKISYEASCNLPESKDPYSILYQKNIEGSETAYYYYVADAETKEID 632
DB 962 FYNESFTNFVYSKADDIN---SLNDESK-----RKLE-----EDIN 995
QY 633 KYRIPQIS---NKNIKLTFIGHGKSEFNTDTPANLDVDSLSSEIETILNLAADISP 688
DB 996 KLKKTQLSPDLNKNYKYLKLERLFDKKTGVKVKQIKLTLKLEQESKLSLN---NP 1052
QY 689 KYIEINLLGCMFYSYSISAE--ET-----YFKULLIKTKDRVSELMPSIQD 733
DB 1053 KHVQNF--SVFPNKKKEAIEAETENTLENTKILLKHYKG--LVKYNGESSPLKTLSE 1108
QY 734 SITVSAN-----QYEVRIEEGK-----REILDHSGKWINKEESI-KD-ISSKEYIS 779
DB 1109 SIQTDENYASLENFKVLSLEGKLDKNLKNLEKKSLSYLSGLHLHIAELKEVKNKNYTG 1168
QY 780 FNKENKIIVKS-----KYLHE---LSTLLQIRNANSSDIDLEKKVMLTEC-----E 825
DB 1169 NSPSENNTDYNNALESYKFKLPEGTVDVATVSE--SGSDTLEOSQPKFASHTVGAESNT 1226

QY 826 INVASNDROI-----VEGRIEEAKNLTSDINVIKNEFKLIESISDLYDKHONGL 878
 Db 1227 ITTSQNVDDVDDVIVIPIGSEEE-----DYDDLQGVVGEATVPSVIDNI-----1273
 QY 879 DGHFISFEDISKETNGFRIRFINKETGNSIFITEKEIFSEYATHISKEISNIKDTI--936
 Db 1274 -----LSKIENEYEVLYLKPLAG--VYRSLKKQLENNVMTF---NVNVRKILNS 1317
 QY 937 -----FDNV-NGKLVKKNLDAAEHV-----NTLNSAFFIOSLIEYNTTKESL-SNLS 982
 Db 1318 RFNKRENFKNVLESDDLIPYKDLTSSNVVVDPKYFLNKKRDKRDLFLSSVNYIKDSIDTDIN 1377
 QY 983 VAMKVQVYAOQFSTGLNTITDASK 1006
 Db 1378 FANDVLGYKILSEKYSKLDLSIK 1401

RESULT 11
 MYS1_YEAST STANDARD; PRT; 1928 AA.
 AC P08964;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin-1 isoform (Type II myosin).
 GN MYO1 OR YHR023W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=91088308; PubMed=2263482;
 RA Sweeney F.P., Watts F.Z., Pocklington M.J., Orr E.,
 RT "The MyoI gene from Saccharomyces cerevisiae: its complete nucleotide
 Nucleic Acids Res. 18:7147-7147(1990).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favellio A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RA "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 VIII.";
 RT Science 265:2077-2082(1994).
 RL [3]
 RN SEQUENCE OF 1-760 FROM N.A.
 RP STRAIN=S288C;
 RC STRAIN=S288C;
 RX MEDLINE=8811539; PubMed=3322809;
 RA Watts F.Z., Shields G., Orr E.;
 RT "The yeast MYO1 gene encoding a myosin-like protein required for cell
 division.";
 RL EMO J. 6:3499-3505(1987).
 CC -!- FUNCTION: Required for cell division.
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X53947; CAA37894.1; --
 DR EMBL; X06187; CAA29550.1; --
 DR EMBL; U10399; AB668872.1; --
 DR PIR; S46773; S46773.
 DR HSSP; P08799; IMND.
 DR GeronOnline; 139340; --
 DR SGD; S0001065; MYO1.
 DR GO; GO:0000142; C:contractile ring (sensu Saccharomycetes); IDA.
 DR GO; GO:0007120; P:axial budding; IMP.
 DR GO; GO:0000910; P:cytokinesis; IMP.
 DR GO; GO:0006970; P:response to osmotic stress; IMP.
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR001609; myosin head.
 DR Pfam; PF00063; myosin_head; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SMO0015; IQ; 1.
 DR SMART; SMO0242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 KW Myosin, Actin-binding; ATP-binding; Coiled coil.
 FT DOMAIN 1 793
 FT DOWAIN 794 823
 FT DOWAIN 856 1911
 FT NP BIND 180 187
 FT DOWAIN 460 529
 FT DOWAIN 36 36
 FT CONFLICT 46 46
 FT CONFLICT 59 59
 FT CONFLICT 86 86
 FT CONFLICT 330 330
 FT CONFLICT 343 343
 FT CONFLICT 421 426
 FT CONFLICT 515 515
 FT CONFLICT 529 535
 FT CONFLICT 541 541
 FT CONFLICT 550 551
 FT CONFLICT 573 573
 FT CONFLICT 582 582
 FT CONFLICT 588 599
 FT CONFLICT 599 599
 FT CONFLICT 627 632
 FT CONFLICT 695 695
 FT CONFLICT 736 742
 FT CONFLICT 756 756
 FT CONFLICT 773 784
 FT CONFLICT 793 794
 FT CONFLICT 896 896
 FT CONFLICT 900 900
 FT CONFLICT 906 906
 FT CONFLICT 911 911
 FT CONFLICT 915 930
 FT CONFLICT 934 939
 FT CONFLICT 951 953
 FT CONFLICT 955 958
 FT CONFLICT 1002 1002
 FT CONFLICT 1049 1049
 FT CONFLICT 1056 1056
 FT CONFLICT 1060 1060
 FT CONFLICT 1085 1085
 FT CONFLICT 1123 1123
 FT CONFLICT 1133 1133
 FT CONFLICT 1144 1146
 FT CONFLICT 1159 1168
 FT CONFLICT 1179 1181
 FT CONFLICT 1184 1185
 FT CONFLICT 1188 1204
 FT CONFLICT 1224 1224
 FT CONFLICT 1228 1228
 FT CONFLICT 1253 1253
 FT CONFLICT 1311 1323

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U67572; AAB99331.1; -

DR PIR; A64465; A64465.

DR TIGR; MJ1322; -

DR HAMAP; MF 00449; -

DR InterPro; IPR003593; AAA ATPase.

DR InterPro; IPR003439; ABC transporter.

DR InterPro; IPR007517; Rad50_zn_hook.

DR InterPro; IPR003405; SMC_C.

DR InterPro; IPR003395; SMC_N.

DR Pfam; PF04423; Rad50_zn_hook; 1.

DR Pfam; PF02483; SMC_C; 1.

DR Pfam; PF02463; SMC_N; 1.

DR ProDom; PD000006; ABC transporter; 1.

DR SMART; SMO0382; AAA; 1.

KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.

FT NP BIND 32 39 ATP (BY SIMILARITY).

FT DOMAIN 158 849 COILED COIL (POTENTIAL).

SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;

Query Match 4.78; Score 242; DB 1; Length 1005;

Best Local Similarity 21.2%; Pred. No. 0.014;

Matches 210; Conservative 165; Mismatches 279; Indels 338; Gaps 59;

QY 4 VNKAQLQKVVYKFRICQEDSWALLNALEB---YHNMGSVVYKYLKLDNNLTQNY 59

DB 132 IKQGEIAKFLSLK-PSEKLTAKLIGIDFEKCYQKQGE--IVKYEKRLERIEGELNY 188

QY 60 LNTYKSGRNKALKKPKFXYLTMTVELELKNLSPVVKNLHFIWIGGQINDTAINYNQWK 119

DB 189 KENYEKELKKK-----MQLEKKNKLMKLEINDKLN--KIKKEFEDIE-KLFNEWE 235

QY 120 DVNSDYTVKVFYDSNAFLINTLTK-----TIVSATNNTLSEFRENLDNPEFD 167

DB 236 N-----KLLYK--FINKLEKRALEKQELKILEYDLNTVVEA-RETLNHKOE 285

QY 168 YNKF-----YKGMELIYKQKHPIDYK--SQIENPEFIDNIKIYLSNEYSKDLE 219

DB 286 YEKYKSLVDIRKTESRLKSHYEDYLTQKLE-----IK-----GDIE 328

QY 220 ALNKVIEESLNKITANNNDIRNLE-----KFADEDLVRLNQVELVERWNLAAASDILRS 275

DB 329 KLEFINKSKYR-----DDINDLTLLNKIKOE-----IERVE 361

QY 276 MKEDGGVLDVILPGIQDLPKSNKPSITNTSMWKLEAIMKYK---EYIPGYTS 332

DB 362 TIK-----DLL-----ELKNLE-----EIEKIEKYKICECKEYVE 395

QY 333 KNFMDLDEVORS-----FESALSSKSKSEIFLDDIKVSPLEVKVAFANNVINQAL 387

DB 396 KYLEEKAEBYKLTLEYITLLQEKKSIEK-----NINDLETRINKLLEETKNIDI 447

QY 388 ISLXDSVCSDLVINQIKRNLNDNLNPSINSGTDNTMTKIFPSDKLASINEDNMFM 447

DB 448 ESIENS-----LKEIEKKVVL-ENLQ-----KXGLGINSE--IKRL 488

QY 448 IKITNYLK--VGPAPDVRSTINLSGPGVYTGAYQDLMLFQKDNSTNHLLEPELRNFEFPK 505

DB 489 KKILDELKEVEGKPLCKTPI-----DENKMLINQH-----K 522

QY 506 TKISQTEQETISLWSFNQ--ARAKSQFEYKKGFGEGALGEDNDLDFQNTVLDKDYVS 563

DB 523 TQLN---NKYTELEENKKIREIEKIEKLKK-----EIDKEENLTKLTLYLEK----- 569

QY 564 KXILSSMKTNRNKEYIHIVQLQGDKISYEASCNLFPSKDYPSSILYQKNIEGSETAYYYV 623

DB 570 -----QSQIEELSILKKNY- 583

QY 624 ADASIKEDKXRIPIQISNKNKNIKLTFTGHGKSEFNTDTFANLVDVDSLSSEISTILNLAK 693

DB 584 -KEQLDEINK-----KLSN-----YVINGK-----PVDEILEDIKSQLNKKF 619

RESULT 13

MSPI_PLAF3
ID MSPI_PLAF3 STANDARD; PRT; 1682 AA.
AC P19598; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMW) (P190).
DE MSP-1.
GN Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5934;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE=88166657; PubMed=3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen precursor p190 of Plasmodium falciparum lacks tripeptide repeats."
RL EMBO J. 6:4137-4142(1987).
RN [2]
RP SEQUENCE OF 1032-1682 FROM N.A.
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1."
RL Exp. Parasitol. 81:47-54(1995).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).
CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42 kDa and 19 kDa antigens which are the major surface antigens of merozoites. The maturation take place during schizont.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M35727; AAA29715.1; -

DR EMBL; Y00087; CAA68280.1; -

DR EMBL; Z35326; CAA84555.1; -

DR InterPro; IPR006209; EGF_like.

DR Pfam; PF00008; EGF; 1.

KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.

KW SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.

FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;
Query Match 4.6%; Score 241; DB 1; Length 1682;
Best Local Similarity 19.7%; Pred. No. 0.028;
Matches 206; Conservative 178; Mismatches 376; Indels 286; Gaps 49;
QY 36 NMSSESVVEKYLKLDIINLTNYKSGSRNKALEKFEYLTMVELEKNNSLTPVE 95
DB 100 NPSDDSSDSDAKSYADLKHVRQNYLFT-----IKELKYPDLFTNTHMLTCD 147
QY 96 KNLHFIWIGQINDTAINY-INQKDVNSDYTVKVFYDSNAFLINTLKTIVTESATNTL 154
DB 148 -NIH-----GPKYLDIGYEIN-----ELLYKLN-FYFDLLBZAKLNDVCANDYC 189
QY 155 E-SFRENLDNPDYNNK-----PYRKRMEIYDQKHFDIYKQSEENPEFIIDNIKTY 209
DB 190 QIPFNLRKANELDVLKLVFGYRPLDFPKDNGKMDYIKKN-----KTT 236
QY 210 LSNVSKDLKALNKYIEESLNKINTANGNDIRNLEK---FADEDLVRLYNOELVERWLA 266
DB 237 IAN-----INELIEGSKTIDQNKADNEGKKUYQAYDLF-IYNQLOBAHNL 287
QY 267 AASDILRIMLKEDGGVLDVILPGIOPDLFKSINKPDSINTSWEMIKLEAMKYKEY 326
DB 288 SVLE-KRITLKNENIKKLLIEDIKIDAEKPTGVNQILS-----LRLEKSRHEK 341
QY 327 IP--GYTSK-NFEDMLDEEVQSFESALSSKSEIFPLDDIKVSPLEVKIAFANNSVI 383
DB 342 IKELAKTIKFNDRUFTD-PLELEYLYREKKNKVDVTPKSDPTKSVQIPKVPYENGIVY 400
QY 384 NQALISLSDSYSDVLVINOIKRY-KILNDNLNPSINEGTDNTMTKIFSDKLASINED 442
DB 401 PLPLTDIHNLAAD-----NDKNSYGLMNPHTKEKINE-----KIITD-----NKE 442
QY 443 NMFMFIKITNLYKVGAPDVRSTINLSGPGVTVGAYQDLMFKDNSTHILHLEPELRNFE 502
DB 443 RKIFI-----NNIKKQIDL-----BEKNIN 462
QY 503 FPKTKISQTEQITSLWGFQARAKSQPEYKKGYFEGALGEDDNLDPAQNTVLDKDYV 562
DB 463 HTYEQNKLL-----DYEKSKDYELLEKEFYEMKFNNFKN-----VVDKIFS 508
QY 563 SKKILSMKTR-NKEYIH-----YIVOLGDKISYEASCNLFSPKDPYSILYQKNIEGSE 616
DB 509 ARYTVNVEQRYNNKFSNNNSVYNQVKLKALSY-----LEDYS-----LRKGISEKD 557
QY 617 TAYVVYV--ADAEIKEDKIPYQISNKRN--IKLTFIG--HGKSEFTNTFANLDV- 668
DB 558 FNHYTLTKTGLEADIKKLTAE-----EIKSENKILEKFNKGLTHSAN-----ASLEVS 605
QY 669 DLSLSSEIETILNKAADISPKYIEINLLGCNMFYSISAEETYPGK-----LLKI 719
DB 606 DIVLQVQVLLIKKIEDLRK-IEFLKNAQL-KUSIHVPNIYKQNKPEPYLVLKKE 663
QY 720 KDRVSELM-----SISQDSITVSANQYE-----VRINEEGREILDHSGK 760
DB 664 VDLKLEFIPKVKMLKKEQAVLSSITQPLVAASETTEDGHSHTHTLSQSGETEVEETEE 723
QY 761 WINKESIKIDISSEKVIYISFNPKENKIIVKSKYLHELSTLLOIRNANSSDIDLEKKVM 820
DB 724 TVGHTTTVT-----ITLPPKEVKVENS-----IEKSDNDSQALTKTVY 763

QY 821 LTECEINVASN-----IDRQIVEGRIEAKNLTS-----DSINYIKNEFKL 861
DB 764 LKLDLBEFLTKSYCHKYLIVSNSSMDQKLELYNTLPBENELKSCDRDLLENQNNIPA 823
QY 862 IESISDSL-YDLKHQNGLDSDHFIISFEDISKTEGFRIRFINKETGNSIFITEKEI--- 917
DB 824 MYSLYSDMNDLQH-----LFFELYQXEMIVYLHLKKEENHIKKLEEPKQITGT 873
QY 918 -----FSEYATHISKEISNIKDTIDFNVNGKLYKK--VNLDAAHEWTLNSAF 963
DB 874 SSTSSFGNTVTNTAQSATHSNQNOOSNASSTNTQNGAVSSGPVAVVEHSDPLTVLSI- 932
QY 964 FIQSLJEYNTTKESLNLSVAMKVOV 989
DB 933 -----SNDLKGIVSLLNLGNKTKV 951
RESULT 14
MSPI PLARC
ID MSP1_PLARC STANDARD; PRT; 1726 AA.
AC P04934;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Merzoite surface protein 1 precursor (Merzoite surface antigens)
DE (EMBL) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5835;
RN [1].
RP SEQUENCE OF 1-1103 FROM N.A.
RX MEDLINE=86205236; PubMed=3517809;
RA Weber J.L., Leininger W.M., Lyon J.A.;
RT "Variation in the gene encoding a major merozoite surface antigen of the human malaria parasite Plasmodium falciparum."
RL Nucleic Acids Res. 16:1206-1206(1988).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).
CC -!- PFM: Merzoite surface antigen contain the sequence of 83 kDa, 42 kDa and 19 kDa antigens which are the major surface antigens of merozoites. The maturation take place during schizont.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; X03831; CAA27446.1; --
DR PIR; A23386; SAZQGM.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merzoite; Polypeptide; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1726 MERZOITE SURFACE PROTEIN 1.
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 SQ SEQUENCE 1726 AA; DBAD45FA352BCF3 CRC64;
 Query Match 4.6%; Score 240; DB 1; Length 1726;
 Best Local Similarity 20.4%; Pred. No. 0.032;
 Matches 211; Conservative 172; Mismatches 361; Indels 292; Gaps 52;
 QY 51 DIANNLTNYNTYKSGRNKALKKFKYELTWEVLEKLNKSLTPVEKNLHPIWIGGQINDT 110
 DB 154 DKHVRVNYLT-----IKELYPELEDTNHEMLTCD-NIH----- 189
 QY 111 AINY-INQKDVNSDYTKVYDSNAFLINTLKKTIVESATNTLE-SFRENLDPEFY 168
 DB 190 GPKLIDGYEIN-----ELLYKLN-FYPDLRAKLVNDVCANDYQIPFNKIRANELD 243
 QY 169 NK-----FYRKRMEIYDKQHFIDYKYSQIEENPEFIIDNIITYLSNEYSKDLKALNKY 224
 DB 244 LKLVFGYRPLDNKDNVGMEDYIKGN-----KTTIAN-----INEL 282
 QY 225 IBSLNTKITANNNGDIRNLEK---FADEDLVRLYNQELVERWNLAAASDILRISMLKEDG 281
 DB 283 IEGSKTIDQKNADNEEGKKLYQAYDL-SIYNKOLEAHNLISVLE-KRITDTLKNE 340
 QY 282 GVLVDVILPGIO-PDLFKSINKPDSITNTSWEMIKLEAMKYKEYIPGYTSKNFOMLDE 340
 DB 341 NIKELLKINEIKNPPANSNGTNTLLDKNKKIEHEE---KIKE-IATKIKENISLFT 397
 QY 341 EVQSFSAJSSKSKSEIFPLDDIKVSPLEKIAFANNSVINOALISKSYCSDLVI 400
 DB 398 D-PLELEYLREKKNKVDVTPKSDQTPKSVQIPKVPYNGIVYPLPLTDIHNSLAAD--- 453
 QY 401 NQIKNRY-KILNDMLNFSINEGTDFNTMKIFSDKLASINEDNMFMKITYNLKVGPA 459
 DB 454 -NDRNSYGDLMNPDTKEKINE-----KIITD-----NKKRIFI-----NNIK----- 490
 QY 460 PDVSTINLSGPGYTGAYQDLMLFKDNTNHLLEPELNFEPFKTKISQLETSQITSL 519
 DB 491 -----KOIDLEKKINHTKEQNKULL 511
 QY 520 WSNFQARAKSOFEBYKGYFEGALGEDDNLDFQNTVLDKDYVSKILSSMKTRNKE--- 576
 DB 512 EDYE--KSKDYBELLEKFE-----MKFNN--FDKQVD-KIFSARYVNVKQR 558
 QY 577 YIH-----YVQLQGDKIYSYASCNLFSDPYSSILYQKNIEGETAYVYV---A 624
 DB 559 YNNKFSSNNVNVQKLKALS-----LEDYS---LRKGISEKDFNHYVTLTKGL 607
 QY 625 DABKEIDKYRIPQISNRY--IKLTFIG--HGKSEFNTDTPANLDV-DLSSEIETIL 679
 DB 608 EADIKKLT-----EIKSSENKILKPKGLTHSAN-----ASLEVYDIVKLQVQVL 655
 QY 680 NLAKADISPKYIEINLLGCMFYSISIAEETYPCK-----LLKIKIDRYSELMP--- 728
 DB 656 LIKKIEDLRK-IEFLKNAQ-KDSIHVPNIYKQNPKEPYLLIVLKEVDKLEKFEIPKV 713
 QY 729 -----SISQSDITSANQYE-----VRINEEGREILDHSGRWINKESITKD 771
 DB 714 KDLMLKEQAVLSSITQPLVAASETTEDGGHSTHTLSQSGTEVTEETEE---TEETV--G 768
 QY 772 ISSKEYISFNPKENKIIVSKYLHELSTLLOEIRNANSDDILEKKVLMTECEINVASN 831
 DB 769 HTTITVITLPPKEVKVENS-----IBKSDNSQALTKTVLKLKLDLEFLTKS 816
 QY 832 -----IDRQIVEGRETEBAKNLTSDSINVIK-----NEFKLIESISDSL-Y 870
 DB 817 YICHKIYIVNSMSMDKLELYNLTPEENELKSCDPLDLFNTQNNIPAMYSYDSMNN 876
 QY 871 DLKHQNGLDSDHSFISFEDISKTENGFIRFRINKETGNSIFETEKEI----- 917

DB 877 DLQH-----LFFELYQKEMIYYLHKLEENHIKLEBEQKQITGTSTSSPGNTT 926
 QY 918 --FSYATHISKEISNIKIDIFDNGVKLVKK--VNLDAAEHVNTLNSAFFIQSLIEVNT 973
 DB 927 VNTAQSAATHSNQOQNASSTNTQNGVAVSSGPAVVEESHDPVLVLSI-----SND 978
 QY 974 TKESLSNLSVAMKVQV 989
 DB 979 LKGIVSLNLGNKTKV 994
 RESULT 15
 MSPI PLAPP STANDARD; PRT; 1726 AA.
 ID MSPI PLAPP STANDARD; PRT; 1726 AA.
 AC PS0455;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMMSA) (GP195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).
 OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89005525; PubMed=3049134;
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
 RA Siddiqui W.A.,
 RT "Plasmodium falciparum: gene structure and hydropathy profile of the
 RT major merozoite surface antigen (gp195) of the Uganda-Palo Alto
 RT isolate."
 RL Exp. Parasitol. 67:1-11(1988).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
 CC kDa and 19 kDa antigens which are the major surface antigens of
 CC merozoites. The maturation take place during schizont.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M37213; AA29611.1; .
 DR InterPro; IPR006209; EGF like.
 DR Pfam; PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 SQ SEQUENCE 1726 AA; 196174 MW; 5B59CEEP2F9A026 CRC64;
 Query Match 4.6%; Score 240; DB 1; Length 1726;
 Best Local Similarity 20.4%; Pred. No. 0.032;
 Matches 211; Conservative 172; Mismatches 361; Indels 292; Gaps 52;

Search completed: April 1, 2004, 16:42:17
Job time : 19.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:38:25 ; Search time 14.25 Seconds
(without alignments)
5885.288 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1020
Perfect score: 5189
Sequence: 1 MNLVNAQLQKVMYVFRQI.....ITDASKVELSTALDETID 1020

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: Pir1:**
- 2: Pir2:**
- 3: Pir3:**
- 4: Pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5189	100.0	2364	2 I40884	cytoxin L - Clostridium sordellii
2	4101	79.0	2366	2 S10317	cytoxin B - Clostridium sordellii
3	4088.5	78.8	2367	2 S70172	cytoxin A - Clostridium sordellii
4	2791	53.8	2710	2 S37052	alpha-toxin - Clostridium sordellii
5	1435.5	27.7	2178	2 S55805	adherence factor T
6	419.5	8.1	3169	2 T00296	adherence factor T
7	350	6.7	3225	2 D81702	adherence factor T
8	348	6.7	2401	2 T28676	adherence factor T
9	345.5	6.7	1127	2 T28317	adherence factor T
10	328	6.3	3724	2 T18427	adherence factor T
11	319.5	6.2	2166	2 G70163	adherence factor T
12	318.5	6.1	3394	2 T18501	adherence factor T
13	316.5	6.1	3255	2 G81702	adherence factor T
14	315	6.1	1302	1 JG6009	adherence factor T
15	315	6.1	2269	2 T28677	adherence factor T
16	306	5.9	3335	2 H81702	adherence factor T
17	303	5.8	2819	2 A90551	adherence factor T
18	294	5.7	1252	2 B42771	adherence factor T
19	290.5	5.6	4688	2 F22885	adherence factor T
20	287.5	5.5	1447	2 F82909	adherence factor T
21	287	5.5	1979	2 C71622	adherence factor T
22	283	5.5	1306	2 T28313	adherence factor T
23	283	5.5	1516	2 E71619	adherence factor T
24	282.5	5.4	2829	2 A42771	adherence factor T
25	280	5.4	1939	2 T18372	adherence factor T
26	272	5.2	980	2 E71606	adherence factor T
27	272	5.2	1711	2 T18429	adherence factor T
28	267.5	5.2	1121	2 F71613	adherence factor T
29	260	5.0	963	2 C90535	adherence factor T

ALIGNMENTS

RESULT 1

I40884
cytoxin L - Clostridium sordellii
C:Species: Clostridium sordellii
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40884
R:Green, G.A.; Schue, V.; Montell, H.
Gene 161, 57-61, 1995
A:Title: Cloning and characterization of the cytoxin L-encoding gene of Clostridium sordellii
A:Reference number: I40884; MUID:95369733; PMID:7642137
A:Accession: I40884
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2364 <RES>
A:Cross-references: EMBL:X82638; NID:G1000694; PIDN:CAA57959.1; PID:G1000695
C:Superfamily: cpl repeat homology
C:Keywords: Cytoxin

Query Match 100.0%; Score 5189; DB 2; Length 2364;
Best Local Similarity 100.0%; Pred. No. 3.6e-188;
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNLVNAQLQKVMYVFRQI	DEYVAI	LNAL	EEYHN	SESSV	VEKYL	KL	KDINN	LDNYL	60
DB	1	MNLVNAQLQKVMYVFRQI	DEYVAI	LNAL	EEYHN	SESSV	VEKYL	KL	KDINN	LDNYL	60
QY	61	NTYKSGRNKALKKFK	EYLT	MEVLE	LKNSL	TPVEK	NLHFI	WIGQI	NDTAIN	YNQWD	120
DB	61	NTYKSGRNKALKKFK	EYLT	MEVLE	LKNSL	TPVEK	NLHFI	WIGQI	NDTAIN	YNQWD	120
QY	121	VNSDYTVKVFYDSNA	FLINT	LKKTIV	ESAT	NNTLES	FRNLND	PEFDY	NKFKR	MEIYY	180
DB	121	VNSDYTVKVFYDSNA	FLINT	LKKTIV	ESAT	NNTLES	FRNLND	PEFDY	NKFKR	MEIYY	180
QY	181	DKQKHFDYKSGQI	ENPEFI	DNIDNI	KYLS	NEYSK	DL	ALNK	YTES	LNKTAN	240
DB	181	DKQKHFDYKSGQI	ENPEFI	DNIDNI	KYLS	NEYSK	DL	ALNK	YTES	LNKTAN	240
QY	241	RNLKFADEDLVRL	YNQEL	VERWNL	AAASD	ILIR	SMLK	EDG	GVLD	VDLP	300
DB	241	RNLKFADEDLVRL	YNQEL	VERWNL	AAASD	ILIR	SMLK	EDG	GVLD	VDLP	300
QY	301	INKPDSITNTSW	EMTKLE	AIMYK	VEYIP	GYTS	KNFDM	LD	DEEY	QSF	360
DB	301	INKPDSITNTSW	EMTKLE	AIMYK	VEYIP	GYTS	KNFDM	LD	DEEY	QSF	360
QY	361	LPDDIKVSPLE	VKIAF	ANNSVIN	QALIS	LKDSY	SD	LVIN	QIKRY	KILND	420
DB	361	LPDDIKVSPLE	VKIAF	ANNSVIN	QALIS	LKDSY	SD	LVIN	QIKRY	KILND	420
QY	421	GTDFTNTMI	FDKSL	ASIN	ENNM	FMKI	TYLKV	GP	ADYR	STIN	480
DB	421	GTDFTNTMI	FDKSL	ASIN	ENNM	FMKI	TYLKV	GP	ADYR	STIN	480

Db 421 GTDFNTTMMKIFSDKLASINEDNMFMKIITNLYKVGPAPDVRSTINLSGPGVYTAYQD 480

Qy 481 LLMFKDNSTNTHLPELPELNFPPKTKISQTEQITSLWSPNQARAKSQFEYKKGYPE 540

Db 481 LLMFKDNSTNTHLPELPELNFPPKTKISQTEQITSLWSPNQARAKSQFEYKKGYPE 540

Qy 541 GALGEDDNLDPQNTVLDKQVSKKILSMKTRNKEYIHYIYVQLQGDKISYEASCNLPSK 600

Db 541 GALGEDDNLDPQNTVLDKQVSKKILSMKTRNKEYIHYIYVQLQGDKISYEASCNLPSK 600

Qy 601 DPYSILYQKNIEGSETAYYYVADAIEIKDRIPIYQISNKNRNIKLTFTIGHGKSEFNT 660

Db 601 DPYSILYQKNIEGSETAYYYVADAIEIKDRIPIYQISNKNRNIKLTFTIGHGKSEFNT 660

Qy 661 DTFANLVDLSSEITETINLAKADISPKYIEINLLGCNMFYSISABETYPGKLLKIK 720

Db 661 DTFANLVDLSSEITETINLAKADISPKYIEINLLGCNMFYSISABETYPGKLLKIK 720

Qy 721 DRVSELMPSISODSLTVSANQVEVRINEEGKREILDHSGKWINKEESIIDKISKEYISF 780

Db 721 DRVSELMPSISODSLTVSANQVEVRINEEGKREILDHSGKWINKEESIIDKISKEYISF 780

Qy 781 NPKENKIIVKSKYLHELSTLLOEIRNANSDDILEKKVMLTECEINVASNIDQIVEGR 840

Db 781 NPKENKIIVKSKYLHELSTLLOEIRNANSDDILEKKVMLTECEINVASNIDQIVEGR 840

Qy 841 IBEAKNLSDSNINYIKNEFKLIESDSLYDLKHONGLDSDHFIKSTENGFRIRF 900

Db 841 IBEAKNLSDSNINYIKNEFKLIESDSLYDLKHONGLDSDHFIKSTENGFRIRF 900

Qy 901 INKETGNSIFITEKEIFESEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEVNTLN 960

Db 901 INKETGNSIFITEKEIFESEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEVNTLN 960

Qy 961 SAFFIQSLIEYNTTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

Db 961 SAFFIQSLIEYNTTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 2

S10317

toxin B - Clostridium difficile

C:Species: Clostridium difficile

C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 15-Oct-1999

C:Accession: S10317; S21894; S22434

R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.

Nucleic Acids Res. 18, 4004, 1990

A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.

A:Reference number: S10317; MUID:90326540; PMID:2374729

A:Accession: S10317

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-2366 <BAR>

A:Cross-references: EMBL:X53138; NID:G40442; PIDN:CAA37298.1; PID:G40443

R:Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.

submitted to the EMBL Data Library, July 1991

A:Description: Comparative analysis of Clostridium difficile toxins A and B.

A:Reference number: S21894

A:Accession: S21894

A:Molecule type: DNA

A:Residues: 1271-2366 <EIC>

A:Cross-references: EMBL:X60984; NID:G40445; PIDN:CAA43299.1; PID:G40446

R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.

Mol. Gen. Genet. 233, 260-268, 1992

A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.

A:Reference number: S22434; MUID:92293124; PMID:1603068

A:Accession: S22434

A:Molecule type: DNA

A:Residues: 1791-2366 <VON>

A:Cross-references: EMBL:X60984

C:Genetics:

A:Gene: toxB

C:Superfamily: cpl repeat homology

C:Keywords: cytotoxin

Query Match 79.0%; Score 4101; DB 2; Length 2366;

Best Local Similarity 77.6%; Pred. No. 3.9e-147;

Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;

Qy 1 MNLVNAQLOKVVYKFRIOEDEYVAIILNALEVEYNNSESSWVKYKLKLDINNLTNDYIL 60

Db 1 MSLVNRKQLEKMANVRFTQDEDEYVAIILDALEVEYNNSESSWVKYKLKLDINSITDIYI 60

Qy 61 NTYKSGRNKALKKFKELYLMEVLELKNLSLTPVEKNLHFIIWGGQINDTAINYNQWKD 120

Db 61 DTYKSGRNKALKKFKELYLMEVLELKNLSLTPVEKNLHFIIWGGQINDTAINYNQWKD 120

Qy 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLSPRENLDPSFDNKNFVKRMEIYI 180

Db 121 VNSDVNVANVFYDSNAFLINTLTKTIVESAINDTLSPRENLDPSFDNKNFVKRMEIYI 180

Qy 181 DKOKFIDYVKQJIEENPEFIIDNIIKTVLSNEYSKDLEALNKYIEESINKITANGNDI 240

Db 181 DKQKNPINTYKQRENPENELIIDIVKTVLSNEYSKDELNTYIEESUNKITQNSGNDV 240

Qy 241 RNLEKFADEDLVRLYNQELVERWNLAASDIILIRISMLKEDGGVYLDVDILPGIQDLPFS 300

Db 241 RNFEBFNGESFNLYEQELVERWNLAASDIILIRISALKEIGMYLDVDMVPGIQDLPFS 300

Qy 301 INKPSITNTSWMTKLEAIMKYKYYI PGYTSKNFMDLDEEVQSFESALSCKSEIF 360

Db 301 IEKPSVTVDWMTKLEAIMKYKYYI PEYTSSEHFDMLDEEVQSFESVLSKSKSEIF 360

Qy 361 LPDDDIKVSPLVKAFANNVINQALISLSDSYCSDLVINQIKRYKILNDNLNPSINE 420

Db 361 SSLGDMASPLVKYAFNKGIIINQGLISVKDSYCSNLIIVKQIENRYKILNLSNLPASE 420

Qy 421 GTDFNTTMMKIFSDKLASINEDNMFMKIITNLYKVGPAPDVRSTINLSGPGVYTAYQD 480

Db 421 DNDFTNTTTFIDSVAEANADGFMELGKYLRVGFPDPVKTITINLSGPPAYAAAQD 480

Qy 481 LLMFKDNSTNTHLPELPELNFPPKTKISQTEQITSLWSPNQARAKSQFEYKKGYPE 540

Db 481 LLMFKEGSNTHLPELPELNFPPKTKISQTEQITSLWSPNQARAKSQFEYKKGYPE 540

Qy 541 GALGEDDNLDPQNTVLDKQVSKKILSMKTRNKEYIHYIYVQLQGDKISYEASCNLPSK 600

Db 541 GSLGEDDNLDPQNTVLDKQVSKKILSMKTRNKEYIHYIYVQLQGDKISYEASCNLPSK 600

Qy 601 DPYSILYQKNIEGSETAYYYVADAIEIKDRIPIYQISNKNRNIKLTFTIGHGKSEFNT 660

Db 601 TPYDSVLFOKNIEDSIAIYYPNGDGEIQEIDKXIPSIISDRPKIKLTFTIGHGKDEFT 660

Qy 661 DTFANLVDLSSEITETINLAKADISPKYIEINLLGCNMFYSISABETYPGKLLKIK 720

Db 661 DTFAGFDVDSLSTEIEAIDLAKEDISPASIEINLLGCNMFYSINVERTYPGKLLKIK 720

Qy 721 DRVSELMPSISODSLTVSANQVEVRINEEGKREILDHSGKWINKEESIIDKISKEYISF 780

Db 721 DKISELMPSISODSLTVSANQVEVRINEEGKREILDHSGKWINKEESIIDKISKEYISF 780

Qy 781 NPKENKIIVKSKYLHELSTLLOEIRNANSDDILEKKVMLTECEINVASNIDQIVEGR 840

Db 781 NPKENKITVSKNLPSELSTLLOEIRNANSDDILEKKVMLTECEINVSITDITQIVEER 840

Qy 841 IBEAKNLSDSNINYIKNEFKLIESDSLYDLKHONGLDSDHFIKSTENGFRIRF 900

Db 841 IBEAKNLSDSNINYIKNEFKLIESDSLYDLKHONGLDSDHFIKSTENGFRIRF 900

Qy 901 INKETGNSIFITEKEIFESEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEVNTLN 960

Db 901 INKETGESIFVETEKIFSEYANHTIEEISKIKGIFDVTNGKLVKKNLDTHEVNTLN 960

Qy 961 SAFFIQSLIEYNTTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

Db 961 AAFPIQSLIEYNSKESLSNLVAMKVQVYAQLFSTGLNTITDAAKVELVSTALDETID 1020

Qy	600	KOPYSSILYKQKNEGSETAYYYVADAEIKETDKYRIPYQISNKRNIKLTFTIGHGKSEFN	659
Db	601	KNPYDSILFQRNIEDSEVAYYNTDSEIQEIDKYRIPDRIPKIKUTFIGHKAEPN	660
Qy	660	TDIFANLDVDSLSSETITINIAKADISPKYIEINLLGCNMFSSYSASBETPGKLLIKI	719
Db	661	TDIFAGLDVDSLSSEETAIGLAKEDISPKSIEINLLGCNMFSSYSVNVEETPGKLLRV	720
Qy	720	KORVSELMPSIGODSTTVSANQYEVREINBEGKREILDHSGKWINKBESIIKOISKEYIS	779
Db	721	KORVSELMPSMSODSIIVSANQYEVREINBEGRRRELLDHSGEWINKBESIIKOISKEYIS	780
Qy	780	FNPKENKIIVKSKYLHELSTLLQEIIRNANSSDIDLEKKVMLPECEINVASNIDROIVEG	839
Db	781	FNPKENKIIVKSNLPFELTLQEIIRNANSSDIELEKKVMLAECINVISNIETQVVEE	840
Qy	840	RIBEAKNLTSDSINYIKNBFKLTIESDSLDYDLHQNGLDSDSHFISFEDISKTEGNFRIR	899
Db	841	RIBEAKNLTSDSINYIKNEFKLTIESI:CDLKQONELEDSDSHFISFEDISDEGFSIR	900
Qy	900	FINKETGNSIFETEKEI:SEVATHISKEISNIKDIFI:PNVAGKLVKKKNLDAAEVNTL	959
Db	901	FINKETGESIFVETEXTI:PFSEYANHITEEISKIKGIFDTPVNGKLVKKKNLDTTHEVNTL	960
Qy	960	NSAFFTQSLIEYNTTKESISNL:SVAMKVQVYAOLFSTGLNTIPDASKVVELVSTALDETI	1019
Db	961	NAAFFTQSLIEYNSKSESISNL:SVAMKVQVYAOLFSTGLNTITDAAKVELVSTALDETI	1020
Qy	1020	D 1020	
Db	1021	D 1021	

RESULT 4

A37052
toxin A - Clostridium difficile
C;Species: Clostridium difficile
C;Date: 31-Jan-1992 #sequence revision 31-Jan-1992 #text_change 24-Sep-1999
C;Accession: A37052; A60991; S21897; S22437; S08638
R;Dove, C.H.; Wang, S.Z.; Price, S.B.; Phelps, C.O.; Lyerly, D.M.; Wilkins, T.D.
Infect. Commun. 58, 480-488, 1990
A;Title: Molecular characterization of the Clostridium difficile toxin A gene.
A;Reference number: A37052; MUID:90129305; PMID:2105276
A;Accession: A37052
A;Molecule type: DNA
A;Residues: 1-2710 <DOV>
A;Cross-references: GB:M30307; NID:G144925; PIDN:AAA3283.1; PID:G144926
R;Wren, B.W.; Clayton, C.L.; Tabacchali, S.
FEMS Microbiol. Lett. 70, 1-6, 1990
A;Title: Nucleotide sequence of Clostridium difficile toxin A gene fragment and
A;Reference number: A60991
A;Accession: A60991
A;Molecule type: DNA
A;Residues: 1' 1894-1899, 'K', 1901-1910, 'Y', 1912-1919; 2054-2074; 2096-2116; 2138-
A;Cross-references: GB:X17194
R;Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sau-
submitted to the EMBL Data Library, July 1991
A;Description: Comparative analysis of Clostridium difficile toxins A and B.
A;Reference number: S21894
A;Accession: S21897
A;Molecule type: DNA
A;Residues: 1-154 <EIC>
A;Cross-references: EMBL:X60984
R;von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.;
Mol. Gen. Genet. 233, 260-268, 1992
A;Title: Comparative sequence analysis of the Clostridium difficile toxins A and
A;Reference number: S22434; MUID:92293124; PMID:1603068
A;Accession: S22437
A;Molecule type: DNA
A;Residues: 1-92 <VON>
A;Cross-references: EMBL:X60984
A;Note: The four fragments shown in reference A60991 correspond to four types o
d with repeats ordered ABCCCDACBDBACCCBACBCDACDCBAC

RESULT 3

S70172

toxin B - Clostridium difficile
C;Species: Clostridium difficile
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 15-Oct-1999
C;Accession: S70172; S44271
R;von Eichel-Streiber, C.; Meyer zu Heringdorf, D.; Habermann, E.; Sartingen, S.
Mol. Microbiol. 17, 313-321, 1995
A;Title: Cloning in on the toxic domain through analysis of a variant Clostridium difficile
A;Reference number: S70172; MUID:96079281; PMID:7494480
A;Accession: S70172
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-2367 <VON>
A;Cross-references: EMBL:Z23277; NID:g761713; PIDN:CAA80815.1; PID:g761714
A;Experimental source: isolate 1470
R;Sartingen, S.; von Eichel-Streiber, C.
Submitted to the EMBL Data Library, July 1993
A;Description: Cloning and sequencing of an Clostridium difficile toxin B mutant.
A;Reference number: S44271
A;Accession: S44271
A;Molecule type: DNA
A;Residues: 1-1323,'N',1325-2367 <SAR>
A;Cross-references: EMBL:Z23277
C;Superfamily: cpl repeat homology
C;Keywords: cytotoxin

Query Match 78.8%; Score 4088.5; DB 2; Length 2367;
Best Local Similarity 77.1%; Pred. No. 1.2e-146;
Matches 787; Conservative 114; Mismatches 119; Indels 1; Gaps 1;

QY 1 MNLVKAQIQKGVYKFRQEDYVAILNALEYHNMSSVVVKYKLDINLTDNYL 60
DB 1 MSVILNKQLEKMANVRFQVEDYVAILDALEYHNMSENVTVKYKLDINSLTDYI 60
QY 61 NTYKSGRNKALKKFKYELTMEVLEKNSLTPEVKNLHFIWIGGOINDTAINYNQWD 120
DB 61 DTYKSGRNKALKKFKYELVIELEKNSLTPEVKNLHFIWIGGOINDTAINYNQWD 120
QY 121 VNSDYTVKVFYDSNAFLINTLTKTIYESATNTLTFSRENLDNPEFDYKFKRKEIYY 180
DB 121 VNSDYVNVFYDSNAFLINTLTKTIIESASNDLTFSRENLDNPEFHHTAFKRWQIYY 180
QY 181 DKQKHFDYKSGQIENPFIIDNIKTLYSNEYKDLKALNKYIIESLNKITTANNNDI 240
DB 181 DKQQNFINTYKAKQENPDIIIDDIKVTLYSNEYKDIIDELNAYIIESLNKVTENSGDV 240
QY 241 RNLEKPADBDVLYNQELVERNLAAADILFISMLKEDGGVYLDVDILPGIQDLEKS 300
DB 241 RNFEEPTGEVFNLYEQESVERVNLGASDILVAIKNIGVYLDVDMLPGIHPDLFKD 300
QY 301 INKPDISI-TNTSWEMIKLRAIMKYKEYIPCYTTSKNFDMLDDEEVQSFESALSSKSDKEI 359
DB 301 INKPDVKTAVDWEQWQLRAIMKKEYIPEYTSKHPDLDDEEVQSFESVLAISKSDKEI 360
QY 360 FFLPDDIKVSPLEVKYAFANNSVINQALISLQDSYCDLVINQIKRYKILNDLNPSIN 419
DB 361 FFLPGLDIEVSPLEVKYAFAGKSIINQALISAKDSYCDLLIKQIONRYKILNDLGLFIS 420
QY 420 EGTDENTWKIKPDKLASTNEQNMFMKITYNLKVGFPADVSTINISGGVYTGAYQ 479
DB 421 QGNDFNTNMNFGESLGAANEENISFIAKIGSLRVGFPEANTTITISGPTIYAGAYK 480
QY 480 DLLMFKDNSTNIHLLPELRNPFPPKTIQSLEQETISLWSFNQARAKSQFEEYKGYF 539
DB 481 DLLLTFKMSIDTSLISSELNRNPFPPKVNISQATEQKNSLWQNEBRAKIQFEEYKGYF 540
QY 540 EGALGEDDNLDPANTVLDKDYVSKILSNKTRNKEYIHYIVQLOGDKIYSASCNLPS 599
DB 541 EGALGEDDNLDPQNTVTDKEYLEKILISSTKSGEGYVHYIVQLOGDKIYSAAACNLFA 600

R:Sauerborn, M.; von Eichel-Streiber, C.
Nucleic Acids Res. 18, 1629-1630, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin A.
A:Reference number: S08637; MUID:90221894; PMID:2109310
A:Accession: S08638
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-55, 'V', 57-2079, 'L', 2081-2549, 'S', 2551-2710 <SAU>
A:Cross-references: EMBL:X51797
C:Genetics:
C:Gene: toxA
C:Superfamily: Clostridium difficile toxin A; cpl repeat homology
C:Keywords: cytotoxin; enterotoxin
F:1820-1839/Domain: cpl repeat homology <CP01>
F:1840-1860/Domain: cpl repeat homology <CP02>
F:1861-1881/Domain: cpl repeat homology <CP03>
F:1933-1952/Domain: cpl repeat homology <CP04>
F:1953-1973/Domain: cpl repeat homology <CP05>
F:1974-1994/Domain: cpl repeat homology <CP06>
F:1995-2015/Domain: cpl repeat homology <CP07>
F:2067-2086/Domain: cpl repeat homology <CP08>
F:2087-2107/Domain: cpl repeat homology <CP09>
F:2108-2128/Domain: cpl repeat homology <CP10>
F:2129-2149/Domain: cpl repeat homology <CP11>
F:2201-2220/Domain: cpl repeat homology <CP12>
F:2221-2241/Domain: cpl repeat homology <CP13>
F:2242-2261/Domain: cpl repeat homology <CP14>
F:2315-2334/Domain: cpl repeat homology <CP15>
F:2335-2355/Domain: cpl repeat homology <CP16>
F:2356-2376/Domain: cpl repeat homology <CP17>
F:2377-2397/Domain: cpl repeat homology <CP18>
F:2449-2468/Domain: cpl repeat homology <CP19>
F:2469-2489/Domain: cpl repeat homology <CP20>
F:2490-2510/Domain: cpl repeat homology <CP21>
F:2562-2581/Domain: cpl repeat homology <CP22>
F:2582-2602/Domain: cpl repeat homology <CP23>
F:2653-2672/Domain: cpl repeat homology <CP24>
F:2673-2694/Domain: cpl repeat homology <CP25>

Query Match 53.8%; Score 2791; DB 2; Length 2710;
Best Local Similarity 52.1%; Pred. No. 1.2e-97;
Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;
QY 1 MNLVNAQLOQVYKFIQDEYVAILNALAEYHNMSESSVVEKYLKLDKNNLTNYL 60
DB 1 NSLISKEELIKLAY-SIRPRENEYKTLTLDYNNKLTANNENKYLQKLKLNESIDVFM 59
QY 61 NTKYKSGRNKALKKPEVLTMEVLEKKNLTPVEKNLHFVIGQINDTAINQWQD 120
DB 60 NKYKTSRRNALSNKQILKEVILKNSNTPVEKNLHFVIGQINDTAINQWQD 119
QY 121 VNSDYTKVYDSNAFLINTLAKTIVESATNNTLESFRENLDPEFDVKNYKRMETIY 180
DB 120 INAEYNIKWDYSEAFVNLTKAIVESTTALQLEEEQNPFQDNKFKYKRMETIY 179
QY 181 DQKHFDYKSCIEBENFEIIDNIKIYLSNEYSKOLEALNKYITEESLANKITANNNGDI 240
DB 180 DQKRFINYKQINQKPTVPTDIDIIKSHLVSEYNRDEVTLESYRTSLRKNNSHGDI 239
QY 241 RNLEKPADDLVRLNQLVSEWNLAAASDIIRISMLKEDGGVYLDVILGIDPDLPKS 300
DB 240 RANSLFTEGELLNYSQELNLRGNLAAASDIIRLLAALNFGGVYLDVILGIDPDLPK 299
QY 301 INKPDITNTSWEMTKLEIMKYKEVIPGYTSKNFMDLDEEVQSFESALSKSKSEIF 360
DB 300 ISRPSIGLDREWEMTKLEIMKYKYNNTSENFDPKLDQQLKDFKLIIESKSEIF 359
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLKDYSQDLVINOIKNRYKILNDNLPISNE 420
DB 360 SKLENLVNSEDLEIKAFALGSVINOALISLQSQSYLTNLVIEQVKNRYQFLNCHLNPATES 419
QY 421 GTDFNTMTKIFSDKLASINEDNMFMKXITNLYKVGFAPOVRSTINISGPGVYTCAYOD 480
DB 421 GTDFNTMTKIFSDKLASINEDNMFMKXITNLYKVGFAPOVRSTINISGPGVYTCAYOD 480

DB 420 DNNFTDTTKIPIHDSLFNSATAENSMTLKIAPIYQVGMPPEARSTISUSGPGVAYASYD 479
QY 481 LLMFKONSTNIHLPELNRNFEFPKTKISQLTQEQITSLWFSNOARAKSQEYKKGFE 540
DB 480 FINQENITEKTKASDLIEFKPPNNLSQLTQEQINSLSWFSNOARAKSQEYKKGFE 539
QY 541 GALGEDDNLDFQANTVLDKDY-VSKILSS--MKTRNKEYIHYIVOLQDQKISYEASNL 597
DB 540 GSLSEDNQGVDFNKNTALDKNYLLNNKIPSNVVEEAGSKNYVHYIIQLQDDOISYEATCNL 599
QY 598 FSKDPYSSILYOKNTEGSETAYYYVAD--AEIKEDIKRIPYQISNKNKILTKIFIGHGK 655
DB 600 FSKNPKNSIIQRNM--NESAKSYFLSDGSEILELNKYRIPERLKNKEKVKVTFIGHGK 657
QY 655 SEFNTDTRANLDOVDSISSEIETILNLAADISPKYIEINLLGCNMFYSISAEETYPGKL 715
DB 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLIDSPKVEVNLGCNMFYSYDFNVEETYPGKL 717
QY 716 LLKIDRVSELMFSPISQDSITVSAQNYEVINEEGREILDSHGKWKINKEESIIKDISEK 775
DB 718 LLSIMDKITSTLPDYNKNSITIGANQYEVINEEGREILDSHGKWKINKEESIIKDISEK 777
QY 776 EYISNPKENKIIVSKYKYLHEJSTLLQEIERNWANSDDIDLEKKVMLTECEINVASNDRQ 835
DB 778 EYIFFSDIKNLKAKSKNIPGLASISDIKTLILDASVSPDTKFLNNLKNIESSIGDY 837
QY 836 IVEGRIEEAKNTSDSINYKNEFKLIESISLSYDLKQNGLDSDHSIFSPEDISKTENG 895
DB 838 IYKLEKLPVKNITHNSIDDLIDEPNLENSDELVELKKNLNDKYLISFEDISKNST 897
QY 896 FRIRFINKETGNSIFETEKEIFSEVATHISKEISNIKDTIPDNVNGKLVKKNLDAAEH 955
DB 898 YSVRFINKSGESVVEVEKEIFSKYSEHITKEITKNSIITDNGNLLDNIQLDHSIQ 957
QY 956 VNTLSAFTIQSLIEYNTTKESLSNLSVAMKQVYVAQLFSTGLNTITDASKVVELVSTAL 1015
DB 958 VNTLSAFTIQSLIDYSSNKVDNLSTSVKQVLAQLFSTGLNTIYDSIQLVNLISNAV 1017
QY 1016 DETID 1020
DB 1018 NDTIN 1022
RESULT 5
S55805
alpha-toxin - Clostridium novyi (ATCC 19402)
C:Species: Clostridium novyi
A:Variety: ATCC 19402
C:Date: 28-Oct-1996 #sequence revision 08-Nov-1996 #text change 05-Nov-1999
C:Accession: S55805; S71294; S71158; S44273; 140834; S44272
R:Hofmann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.
Mol. Gen. Genet. 247, 670-679, 1995
A:Title: Sequencing and analysis of the gene encoding the alpha-toxin of Clostridium no
A:Reference number: 140834, MUID:95342160, PMID:7616958
A:Accession: S55805
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2178 <HOF>
A:Cross-references: EMBL:Z48636; NID:G728537; PIDN:CAA88565.1; PID:G755724
A:Accession: S71294
A:Molecule type: protein
A:Residues: 1-15 <HOF>
R:Hofmann, F.
submitted to the EMBL Data Library, March 1995
A:Reference number: S71158
A:Accession: S71158
A:Molecule type: DNA
A:Residues: 1-1179, 'LKV', 1183, 'LVTHIGE', 1191-2178 <HOF>
A:Cross-references: EMBL:Z48636; NID:G728537; PIDN:CAA88565.1; PID:G755724
R:Hofmann, F.; Habermann, E.; von Eichel-Streiber, C.
submitted to the EMBL Data Library, July 1993
A:Description: Sequence analysis of Clostridium novyi alpha-toxin: a member of the fami
A:Reference number: S44272

Qy 872 LKHONGLDGSHFISFPEDISKTENGFRIRFNKETSIFITEKEBSEVATHISKEISN 931
:
Db 878 FYCKNNISNMWILFSDIIEKDYNNVKLANKITGETSVIKTYSDSLWNFTNKYKKIVDD 937
:
Qy 932 IKOTIPDNVNGKLKVYNLDAAEHVNTLSNAFFIOSLIEYNTTKESLSNISSVMKVOYYA 991
:
Db 938 IKGIIKVDINGEFTFKADFIENQFNSLNSAMLMLQLDIDYKYTEILTNNMNTSLUKOAYA 997
:
Qy 992 QLFSTGLNTITDASKVVVELSTALD 1016
:
Db 998 QIFOLSIGAIOEATEIVIITSALN 1022
:

RESULT 6
toxin B - Escherichia coli plasmid p0157
C/Species: Escherichia coli
C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 17-Nov-2000
C/Accession: T00296; T42195
R/Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota,
S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A/Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic
E.coli O157:H7 strain, EDL933, MUID:98290540; PMID:9628576
A/Reference number: Z14127; MUID:98290540; PMID:9628576
A/Accession: T00296
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3169 <WAK>
A/Cross-references: EMBL:AB011549; NID:G4589740; PIDN:EAA31815.1; PID:G3337056
A/Experimental source: strain EHEC O157:H7, substrain RIMD 0509952
R/Bertrand, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A/Title: The complete DNA sequence and analysis of the large virulence plasmid of Esche-
ria coli O157:H7 strain EDL933, serotype O157:H7
A/Reference number: Z22068; MUID:98391744; PMID:9722640
A/Accession: T42195
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-236,'D',238-1887,'I',1889-3169 <BUR>
A/Cross-references: EMBL:AF074613; PIDN:AAC70163.1
A/Experimental source: strain EDL933, serotype O157:H7
C/Genetics:
A/Gene: toxB
A/Genome: plasmid p0157
A/Note: L7895
C/Superfamily: Escherichia coli probable cytotoxin
C/Keywords: Cytotoxin

Query Match 8.1%; Score 419.5; DB 2; Length 3169;
Best Local Similarity 20.8%; Pred. No. 4.1e-08;
Matches 256; Conservative 202; Mismatches 446; Indels 325; Gaps 55;

Qy 14 YVKRI---QEDEVAILNAL--FEYHNMSESSVVEKYLKDKDINLTNDYINTYKSGR 68
:
Db 160 YIKRKIRGAEDQTITTQSIIINELNGVDNRITI-PFOKISELNDIHVSVENMOIKNSR 218
:
Qy 69 -----NKALKKFKE----YLTMEVLELKNNSLATPEKNLHFITCG 105
:
Db 219 KGIBLVKQGELLSSLINVNKGNQLSDNASKIINLLGIEYQSHKV-DIEPFIHAVVAG 277
:
Qy 106 QINDTAINYINQWKVNSDYTVKVFYDSNAF-----LINTLKT----- 144
:
Db 278 APPDNTFSYITAFNLTKDYTYLLWDPNFAAKFGSKLNIAAMVAIMELRITNPFLA 337
:
Qy 145 -----IVESATNWILE--SFRENLDPEFDY-----NKEYKRMELIIDYKQKHF 186
:
Db 338 EEMNEVLKIQIQNETIEFKETRIKLENRYKSITSETKEKPNVFLESIMGMQDNY 397
:
Qy 187 IDYKKSQIEENPEPI----IDNIHK--TYSNVEYSKDEALNKYIEBSLNKITANNNG- 238
:
Db 338 FTYCISNCISNTDDISRDLDFTNVLKUSPFVQNDFKSTVEKNKRDIIDLKWTISOQFGDR 457
:
Qy 239 ----DIRNLEKFADEDIVRLYNQBELVERWNLAASDILRI-SMLREDGGVYLDVDPGIG 294
: :

Db 458 PQLRDINTLSFFKKPDQFFYQOEMLRWNYAASDQVRINIILKEGGIYTDITLIPAYS 517
Qy 295 PDLFSKNKPSINTNTSWEMIKLEAMKYK--EYIPG--YTSKNFMDLDEVQSRFESAL 350
Db 518 DKVQOLINE-KSDDKRRFFEDKLRIISSESILSLIKEKYSIKH-DGLDDBTTINQLNNIL 575
Qy 351 SSKSDKSEIFLDDDIKVSPLFVLEKIAFANNVSI-----NOALISLKDSY 394
Db 576 -SEIEK---LTIDDY-PKPVETKVRVDTFKIPKRYQKWTENTWIRGNNNFMTLHGSK 629
Qy 395 CSDLVINQIKNRYKILNDNLNPSINEGDFNTTKIFSDKLASINB-----DN 443
Db 630 CIDFILSQKQXLEYL-QRINDNTSYNNLFYTT-----EDLKSINNVAIGGIPAKKYLEH 683
Qy 444 MFMWIKITNYLKVGFAFDVRSNTINLSCP-----
Db 684 GLP-----SEYRQDGTIPYVSTNLISGPDIMRQMKYKSLGRIGBHVHKDKNLSVNF 739
Qy 472 -GVYTAGVODLLMF---KDNSTNTHLEPE-----LRNFEPPKTKISQLT---EQEIT 517
Db 740 LOVTASSNKDNKSNFNLNPSVGVINDITPDDESSWAVRNNDINKILPEKINCHVPEKLPT 799
Qy 518 SLWFSNQARAKSQPEEYKKGFEFEGALGEDDNLDPQNTVLDKQYVSKKILSSMKTRNKEY 577
Db 800 SLY-----YEIDSRSPFGW-----DNKSIKHVTETNKDILI--KDNILLTSSNID 843
Qy 578 IHYIVOLGDKISYEASCNLFSDPKDYPYSILYQKNIESEYATYYIVADAKIIDEKTRIP 637
Db 844 VKLLIKL--DRELVASISKI-----DNPLALRSIRTLQLQLANVYTSNTPEPENTINFI 895
Qy 638 YQISNKN-----IKLTFIGHGKSEFNTDFANLVDVS-----LSSEIE 676
Db 896 YDFVRKQDQDLSAFL-----FSNADTKLIWVNSMKNVFLREVIS 941
Qy 677 TILNLAADISPKEYIENLLGNMFSYSISAEETYPGKLLIKIDRVSELMPSISQDSI- 735
Db 942 CVLRSKKVD---SYINEN-----KNLSKEDA--GALRDYAKLXMKELFSMLDDDGK 989
Qy 736 -TVSANQYEVRAINEGKEILDHGKWKNKESIKDISKEYISFNP-----782
Db 990 KIITTNAY---IKERDKL-----SGIIYNIENSIHGESPDITRSNHEWGDLSLVEQF 1041
Qy 793 KENKIIVSKYILHELSLTLQIRNANSSDIDLEKKWMLTECEINVAS-----NIDQIV 837
Db 1042 KKFEPYKSE-LSSAKSFDDIKNK-YITDPETKRNVLHQLSDIKERIAFLDISHAY 1099
Qy 838 EGRIEAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHONG 877
Db 1100 FGSLEKLQLS-----GYVFSINIIEYLLASGVSGHYGVYVPAPSKLLELRHT 1155
Qy 878 LDDSHFISFEDISTENGFRIRINKETGNSI---FIETEKEIFSEYATHISKEISNID 934
Db 1156 KSNSEWI--EKITP---YVVDILSDNSVNLRPPLSEEQKILNDIKLEISKVS---E 1206
Qy 935 TIFDNNVGNKLVKNLDAHEVNTLNSAFFTCQSLIEYNTTK-----ESLSNLSV---983
Db 1207 QYFMKLTQKSSVIGIKYSVDPDRYNENLFLSLPINQNLTLFFMYRYPEMLYDIHIGIE 1266
Qy 984 --AMKVQVYQALFSTGLNTITDASKVVEL 1010
Db 1267 NKANREFIYKFSNLNLDLINDERVLNL 1295

RESULT 7
DB1702
adherence factor TC0439 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: DB1702
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000

A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: DB1702
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3225 <TEXT>
A:Cross-references: GB:AB002312; GB:AB002160; NID:g7190482; PIDN:AAF39293.1; PID:g719048
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0439

Query Match 6.7%; Score 350; DB 2; Length 3225;

Best Local Similarity 19.7%; Pred. No. 1.8e-05;

Matches 255; Conservative 207; Mismatches 406; Indels 428; Gaps 64;

Qy 2 NLNVAQKQKVVYKFRIOEYVAIL-NALIEEYHNMSE-----SSVVEKYLKLDINN 54

Db 196 NILEKLSITQ-----EQIKHNLNLSNKEALNRSDLNKNSRKQSLLVQAOKILD--E 249

Qy 55 LTDNYLNTYKSGRNKALKFKFYLTMEVLEKNNSLTPPEKNLHFIIWGGCINDTAINY 114

Db 250 ILSQTKSTEERASNSVMTTIKKEFTSHRV-----PVEKNIHGIWAGSPPEGTDEY 300

Qy 115 INQWKNVSDYTVKVYDSNAF-----LNTLKKTIVESATNTLTSPENLNDPFDYNK 170

Db 301 IKLFLHTYPEFSFLPMDKTAYGAAKFSSTLKRIAFDAVN-----SLREATPEPVQFVQ 356

Qy 171 FYRKEWELIYDKQKFFIDYKSOITEENPEFIIDNIIKTYLSNEYSKDL-----218

Db 357 RYDK-LKKSYSRTSDFDE--KORUSEQVELYDNY-----NKFSKETQSNFVLLHEM 407

Qy 219 -----EALNKY-----IEESLNKI 232

Db 408 ITIQSPFNYCOLKVGAITDETREIYLEKVLKVEEDLSHYKETIKNKESIEKLVEKI 467

Qy 233 TANG-----NDIRNLEKFADEDIARLYN--QELVERWNLAASDIILRIMLKEDGGVY 284

Db 468 NDSTGRERVVIKDIRLAKSL--QDLTNSYNYETEMLLANVYAAATDQLRMYMLKEYGGIY 525

Qy 285 LDVDILPGIOPDLFKSN-----KPSITNT-SWEMIKL-----EAIMKYKEVIPGT 331

Db 526 TDLDMPOYSQDVLQKIMDVGSGRFFEDHKLRTLSFAALKLGSGKQTTVSEEAKNAMT 585

Qy 332 SKNFMDLDE-EVQRFESALSSKSKSIFLPLDDIKV---SPLEVKI---AFANNSVI 383

Db 586 LPTFTLQDQKSEIFKYLETQAKSLFQPMQVTVVRDNPILQRHKTGWGTGNVRL 645

Qy 384 NOALISLKDSCYSDVINOIKNRYKILNDNLNPSINEGTFDNTTKIFSDKLASINEDN 443

Db 646 NGLMMAHKDSAVDAVIARQAAAYDEMR-ALRQNVVSGEFFRSL-----GDLEHVNREKN 699

Qy 444 M-MFMIKITNYL-----KVGFAPDVRSNTINLSGPGVYTGAYODLLM--FKDNSTNI 491

Db 700 IGGYLAKE--NYLGGSLFPDFRQDSVIPGAISTLGISGP-----DIIMDTMSDYFTNL 749

Qy 492 -----HLLEPELN-----PEFPKTKISQLT-----EQEIT---SLMSFNOA 525

Db 750 GPVGEDFLYEGKLGKAAFLGAYCAQKTPKGELTYDNLHPLSLGANDVTPADASTCETKQ 809

Qy 526 RAKSQF-----EYKKGYPEGALGEDD-----NLDFAQ-----NTVLDK 559

Db 810 HCAEELLSDSISSEDEHPKGIERRVNPNDPFSKLSKKAQGLISDFADLLPRFNLLIES 869

Qy 560 DYVSKKILSSMKTRNKEYIHYIVQLGDKISYEASCNLFSDPKDYPYSILYQKNIESETAY 619

Db 870 SALDIHTLSAL--DRDIQHLFTKVQ-----KDPVASVA-----900

Qy 620 YYYVADAETKEIDKYR-IPYQISN-----KKNIKLTFIGHGKSEF---558

Db 901 ---VFSLQQLAEMIRAIPEFFIRNOVHILPEAQAHFEADWKKAIQIYLHSHFQTEVINY 957

Qy 659 ---NTDTFANLVDVSLSSSIEIILNLAADISP-----KYIEINLGC-----NMF 701

Db 958 SSTHTQIVFGKOLLAVARAAKSL-MSDHPSLTISYLYKKTQSHLGLVLTFFQEDFF 1016
 QY 702 SYSIS-ABEYTPGKLLKIKDRVSELMPSISQDSITVSANQVEVINEEGKREILDHS-G 759
 Db 1017 ELMDVIAEPELHKLKIEQVN-----SGLYSHVEHSLG 1052
 QY 760 KW--INKEE-----SIKDISKEYISFNPXENKIIVKSKY-----LHEL 797
 Db 1053 EMLKLSKERKSKFLILAKETFEERERDSQQOHTWFEELYEKHQRVQKDPACKIOEL 1112
 QY 798 STLLQIRNANSSDID-----LEKKVMTCEINVASNIDROIVEGRIBEAKNLTSD 850
 Db 1113 ITVFQE-SORVOAQDIDTFAHKPFYQDLMKOGYAFEDISVITKYL-----LASD 1161
 QY 851 SINVIKNEFKLIESDSLYD-LKHONGLDSDHFSFEDISXTENGFRIRF--INKETGN 907
 Db 1162 GVGSIITDPIPPPKQIDAMKQSLG-----EDFGSLHYTLQVYDWLSKET-N 1211
 QY 908 SIFPIETEK-----EIFSEYATH-----ISKEISNIKDTIFDNV-- 940
 Db 1212 SVTSEQAKQKLPKLEHLEGVYTHDLLIPPIDGVSVALGLAFSTEEGKVSDEVLTSTAP 1271
 QY 941 -----NGKLIVKVNLDAAHEVNTLNSAFFIOSLIBYNTTK 975
 Db 1272 GVPNSASYAMTSYLYGLFLITKDIQSGRLTHEI---LKERLQTYGGAYFI-----N 1319
 QY 976 ELSNLISVAM--KVQVY---AQLFSTGLNTITDAS 1005
 Db 1320 ESKIDVLLALSRKKAQISLIDAKHALTGFSPFSEAS 1355
 RESULT 8
 T28676
 rhostry protein - Plasmodium yoelii (fragment)
 C;Species: Plasmodium yoelii
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
 C;Accession: T28676; A45521
 R;Sinha, X.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
 Mol. Biochem. Parasitol. 76, 329-332, 1996
 A;Title: Comparison of two members of a multigene family coding for high-molecular mass
 A;Reference number: Z20507; MUID:97077455; PMID:8920022
 A;Accession: T28676
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-2401 <SIN>
 A;Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1
 R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
 Mol. Biochem. Parasitol. 42, 241-246, 1990
 A;Title: Identification of the gene for a Plasmodium yoelii rhostry protein. Multiple co
 A;Reference number: A45521; MUID:91101660; PMID:2270106
 A;Accession: A45521
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 2260-2401 <KEE>
 A;Cross-references: GB:M34281
 Query Match 6.7%; Score 348; DB 2; Length 2401;
 Best Local Similarity 20.9%; Pred. No. 1.4e-05;
 Matches 262; Conservative 207; Mismatches 406; Indels 377; Gaps 66;
 QY 6 KQALQGVVYKPRIQSDVEVAINALBEEYHN--MSESVVEKYLKLDKNNLTNDYNTY 63
 Db 118 KDTMKKIIILIIQYM-NEFGKNDAMTKLNEGSGISQKFINNQIKQKFKSTYDE----- 171
 QY 64 KSGRNKALKKFK--EYLTWEVL-ELKNNSLTFVEKNLHFVIGQINDTAINYNQKD 120
 Db 172 KKEGFESSLELAKNWEKKLEIITELKKNBEETVQDIREIKIQKD-----IIEEQKI 227
 QY 121 VN-----SDYTVKVFYDSNAPLINTLTKTIVESATNTLBSFRENLDNPFQYDK 170
 Db 228 VNDLKLLELNKKIKETIEYIKAV---DLKKEIKD--NVVID---ELAKEPPYQITK 279
 QY 171 FYRKRMEIIVDKQHFIDYKQSIEE--NPEFII---DNI-----IKTLYLSNEY 214

Db 280 YIEKNEIYNTIKSDFDKIVGDIQLYNMFVQESNIEHTEILTKTKIDNVY 339
 QY 215 S-----KDEALNKVIEESLNKITANN-----NDIRNLEK----- 245
 Db 340 NIQNMETETVSHLKNITNNKLSSETLIDIIKIYGEITNELNKLTEDPKNEKGLSNK 399
 QY 246 ---PADBDL-VRLYNQBLVE---RWNLAAASDILRISMLKE-DGGVYLD--VILPGIQP 295
 Db 400 IDEVAKENVQVLYKSNILEIKKHYN-----DQINIDNIKEAKAQNYDQKEHMTIPP 454
 QY 296 DLFSKINPDSITNTSWEMIKLEAIMKYEYIPGYTSKNFMDLDEEVQSFESALSKSD 355
 Db 455 NEMK-YQKP---SIEIKMKDEFLSKNKY-----NDFKVKYKVESEHNKFTELTN 503
 QY 356 KSEIFLDDIKVSPLEVKIAPANNVINOALLSLKDSY-----C--SDL 398
 Db 504 K--IKTEVSDEEIKKYENKEN-DSKSLINETKKSIEEYQNIINTLKKVDDYIIVCLNTNE 560
 QY 399 VINQIKRYKILNDNLNPSINEGTDENTTMKIFSDKLASI----- 438
 Db 561 LITNCHNRQTTLKDLNQNKTIKETNSIDKIITDKFENILTDKKTETEFTKFTGLSNNH 620
 QY 439 -----SNEDNMKFM-----IKITNLYKVGFPADVRSTINL 468
 Db 621 ESNNKELLTVFYDLKANLGNKNENMLYKQFNEKEKAVEDIKKNV-----DINKIVSN 673
 QY 469 SGPQVYTGAYQDILMFKNSTNIHLLEPELRNEFPFKTKISQLTQEBITSLMWFNOARAK 528
 Db 674 IEITIVTSIYN--INEDTENEIG-----KSTELLNTKYLEKVKANVTNL----- 715
 QY 529 SQPEEYKKGYPGALGEDDNLDFQAQNTVL--DKDYVSKKILSSMKT-----RNKEIYHYI 581
 Db 716 NEIKBKLDYDFQDFGKERNIKYPDENKIKNDITLNQKIDKISFETLTKKXSE--NHI 773
 QY 582 VOLQG--DKISYEASCNLPSKVPSSLYQKNTEGSETAYYYVADAIEKIDK--YRIP 637
 Db 774 DEIKGQIDKLVKVPKNTFNEPD-----KEIEKKIENIV 807
 QY 638 YQISNKRNIKLTPIGHGKSEFNTDTFANLDVDSLSBIEITILNLAADISPKYIEINL-- 695
 Db 808 EKIDKKNI-----YKEIDKLLNEISKIEN-DKTSLE-KLKNINLSY 847
 QY 696 ---LGCNMF-----SYSISABEYTPGKL-LLKIKDRVSELMPSISQD----- 733
 Db 848 GKSIG-NLFLQQIDEEKKAEHTIKAMEAYIDDLNIIKKKSQBIKEMNINMDIKMDIKH 906
 QY 734 ---SITVSANQYEV-----RINEEGKREILDHSGKIN--KEESIIDKDISK----- 775
 Db 907 EMKALNISHDDYKIYHTTSKXHEKISDIRKNSLKIIQDFSEESIINDIKKELEKNVLES 966
 QY 776 -----EYIS-----FN-PKENKIIVKSKYLHELSTLLQEIERN-----ANSSDI 813
 Db 967 QNNTDINQVLSKIENIYNTLKNKI---KKIIDKVEYDEIEKNNKINAELSSEKI 1023
 QY 814 --DLEKKVMTCEINVASNIDROIVEGRIBEAKNLTSDSIN-----YIK----- 856
 Db 1024 ITOLKNESSLKEQSKIKSIIDNYYVSECIKNITNLKTYIVNEKNINITYFKNAEYQN 1083
 QY 857 ---NEFKLIESISDLYLKHONGLD--DSHFISFEDISKTEGNGFRIRFINKETG-NS 908
 Db 1084 VSLNFNNEIADTKSYILNKKNGNTNNTDYNIKELKEHKKSKSNVYK-----DEAGKNT 1138
 QY 909 IFIETEKEIFESEYATHI-----SKEISNIKDTIFDNVNGKLVKVNLDAAHEVNT-- 958
 Db 1139 QEIKONKELFEKEQEVTVLNNKYAVELKNKFDKT-KNVSEQIIEIK--DAHNTFTSQ 1195
 QY 959 -----LNSAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITD 1003
 Db 1196 ADKSEKQWBIKNEQFIEDEVAKNKSNKAI-LDIQLSVEPFIKFLKID 1246
 RESULT 9

T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28317
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:g4049717
C:Genetics:
A:Note: MSV156

Query Match 6.7%; Score 345.5; DB 2; Length 1127;
Best Local Similarity 21.9%; Pred. No. 6.8e-06;
Matches 248; Conservative 167; Mismatches 368; Indels 349; Gaps 54;

QY 9 LQKVVYKFLQDEYVAIINALBEYHNMSESSVVEKYLKLDINLNTDNLTYKKS-- 66
DB 61 VEKIFYMHENQFKINDYNIILQYLIEYNNKICKENKFFCK--NPL---YNTYKCKLY 115

QY 67 --GRNKALKKPKVLTWEVLELKNLSLTPVEKNLHFIWIGGQINDTAINYINQKDVNSD 124
DB 116 IYLDYBEKKDKELVIN--TEQKN-----AVDK-----INDKNNVNIHSNETI 159

QY 125 YTVKVPYDSNAFLNTLTKTIVESATNTLESFRENLDPEFYNNFYKRMELIYDKQK 184
DB 160 ITGK--ETLIDILNKLK-LVSDDEKQLIEQYKNNKIEIFKN----- 201

QY 185 HFIDYKSOIEENPEFIDNIKTYSNEYSKOLEALNKYIBESLNKIKITANNQNDIRNLE 244
DB 202 -----IDNVQK-----EINKQDELNKLKLSKKEFFIKQ-----EELN 235

QY 245 KPADEDIVLYNQBELVERNLAAASDILRISMLKEDGGVLDVDILPGIOPDLFKSINKP 304
DB 236 KTIDKK-----QEBLILKLN-----DKEINFNID-----EKQKLDQI 268

QY 305 DSINTSWMTKLEAIKMK-VKEYIPGYTSKNFMDLDEEVQSFESALSSKSDSEIFLPL 363
DB 269 NSKINTLNENIK--GVNMLYTE-----TKNKSINLQNEILNK-DSTIKSLDEKQLDEL 320

QY 364 DDIKVSPLEVKIAPANNVIN-QALI--SLKDSYCSDLVINQKNRYKI-----LNQN 413
DB 321 DK-NINNTSYLNKSNKTIQNIQQLLESSLTDFNANINELKSKIKLFDNDIQKLND 379

QY 414 LNPISNGTD-FNTMKIPSDKLAS-----TSN-----EDNMFMWIKITNLYKVGA 459
DB 380 ITEQNNKTIQDFNNSTRIFKELDTVEYKIDDIKNNNLQKLESYKKIDQETYYKNKIN 439

QY 460 PDVASTINLSPGV-----YTGAQDILMFKNSTNIHLPELR 499
DB 440 KEYNDIIEELKNNLQKLEENKKIDQETYYKNKINKEYNDIIEELKN--NLOKLEENK 497

QY 500 NPEFPKTKISLTQETITSLMSFNO-----ARAKSOFBEYKIGYEGALGED 546
DB 498 NINKLTKLNDIESN-TEL--FNKLNISDFKDSRIAKLNTYEQLRKOLLENKTN 554

QY 547 DNLDFQNTV--LDKDYVSK-----KILSSMKTRNKYIHYIVOLQDKISYEASCN 596
DB 555 ELMKLSDNKLSLEQLYDSKKNLDGIDKYNLSKEKN-----DKI----- 595

QY 597 LFSKDPYSSILYQKNIEGSETAYYYVADAEIKEDIKRYPYQISNKNKILFIHGKS 656
DB 596 ----DEYFS-----NIEKFDIYNVIENK-----FIGNLD 621

QY 657 ----EFNTDTF--ANLQDVSLSSEIETILNIAKADISPKYIEINLAGQMFYSISABE 709
DB 622 IINKIINDQPKYINKSDSKSNELSTMFD-----DIFNAKNQIASITNNIENTIS 672

QY 710 TYPGKLLIKIKORVSELMPISQDSITVSANQVEVRINERKGRBILDHSGKWINKERSII 769
DB 673 -----NKIKD-----LNERTIS-NEDSSKELLDEIRKY-KQQPDKI 706

QY 770 KDISKEYISPNKENKII--VSKYLHELSTLLQETIRNNANSSDIDLEKKWMLTECEI- 826
DB 707 KDAWNTVEKSFENTLQKDIDSIKSN-INELTNAYDIINTKAN-----DLDKLNAGSFEK 761

QY 827 --NVASNIDRQIVEGRIEBAKNT-----SDSINYIKNEF--KLIE-----SISDS 868
DB 762 NLYNASDILLDTIQNNDEKVKQLNVEYKKNQKQSIINDIVNNFIKELAKFNNTENKS 821

QY 869 LYDLKHQGLDSDHFISEFEDISKTEGFRIRFINKETGN-----SIFETEKEIFS---- 919
DB 822 LNELLTNDINDIKFYLKYLKELNKISTNNLLIKYKNEIDNVNKLISVIENTLQFINSELSI 881

QY 920 -----EVATHIS-----KEISNIKDTIFDNVNGKIVKVV 948
DB 882 EFNQGSITSHINFLNTLAGINDVNLKLNKIMADTTRRGDTNIRDEIKQISSENKISQ 941

QY 949 NLDAAEHVNTLSNAFFIQSLIEYNTT-----KESLSNLSVAMKVQVY 990
DB 942 KFNKNEKDKLKLISFNDKLNKYNISAGYTEYNNIEHECLKYLIVASDQEY 993

RESULT 10
T18427
Hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
Submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:Z98547; NID:el325376; PID:el325379; PIDN:CAB11104.1
C:Genetics:
A:Introns: 307/1; 1545/2
A:Note: C0335c

Query Match 6.3%; Score 328; DB 2; Length 3724;
Best Local Similarity 21.9%; Pred. No. 0.000147;
Matches 230; Conservative 173; Mismatches 347; Indels 302; Gaps 57;

QY 20 QDEYVAIINA-----LEBYHNMSESSVVEKYLKLDINLNTDNLTYKKS-----G 67
DB 618 EDEETIHLKKNLKDANETNNDKEN-----KNWTKELK-SKNYLENEKGTLELKLGR 672

QY 68 RNKALKKPKVLTWEVLELKNLSLTPVEKNLHFIWIGGQINDTAINYINQKDVNSDYTV 127
DB 673 KNNIFKDEKY-----NSLGEV-----IINEIQIENKINDIQDG-- 708

QY 128 KVFYDSNAFLNTLTKTIVESA--TNNTLESFRENLDPEFYNNFYKRMELIYDKQKH 185
DB 709 -----NISKQKIIQSSSRNTDFTNFKDISLAND---DLEKERRK-----KSQH 748

QY 186 FIDYY-----KSQIENPEFIIDNIKTY-----LSNEXSKDL-----EALNKYIBESLNKI 232
DB 749 FIDNLVADKQNEISEINIKICDNNINNYDESINNIYDESINNIYDESINNI 808

QY 233 TANNNDI--RNLEKFADEDLVRLYNQ-----ELVERNLAAASDILRISMLKEDGG 282
DB 809 YDENINNIYDENINNIYDENINNIYDEGINIKICDNNILENKNIKTNDIVQV---EENN 864

QY 283 VYLDVDILPGIOPDLFKSINKPDSITNTSWEMIKLEALMKYKEIPIGVTNKNFMDLDEV 342
DB 865 ESIE-----KNELMISLNKO---INNTYNNFK-----ENVDIFINKI 898

QY 343 QRSFESALSSKSDKSEIFLDDIKVPLEVKIAPANNVINQALISLKSQYCSDLVINQ 402


```

Best Local Similarity 19.1%; Pred. No. 0.00033;
Matches 249; Conservative 194; Mismatches 443; Indels 419; Gaps 57;

Qy      3 LVNAQOLQKVVYKFKRIQEDVEYAILNALAEYENHMFESSVVEKYLKLOINNLNTLYLNT 62
Db      200 IARLKLTBEIQKCRISLNN-----LKAQYQYINIDESH--RNSRKQDULLVKQSIFLNN 254

Qy      63 YKSGRNKALKKFEYLTMEVLBKQNSLTPVEKNLHFIIWIGQINDTAINYNQKVDN 122
Db      255 ILENAMWESBEPKVMVSIKTEF--NSHRVKIDSFHGIWAGAPPETGDDVIKTFLOTY 313

Qy      123 SDYTVKVFYDSNAFLINTLKKTVESATNNTLESFRENLDPEFNKFF----- 172
Db      314 EEPDFLFWDRQAFGAAKFSILKIAFDSSKELRSITSETQDFVEKVEDLQAKYQTS 373

Qy      173 -----RKRMEIYDKKHFD-----YVKSQIEENPEP-----IID 203
Db      374 GRREERLEKDKURKULFDRYKHLUSQEIIRTNFDFALFKNMVLSODGFNYCLLKIGITIDD 433

Qy      204 NIITVLSNYSKDLEALNKY-----IBESLNKITANNND-----IRNLEKFA 247
Db      434 QTRIAVLQDELHLPESIEIQYKDLIKANEQIKDIVGKNTDILGKRVFIKINRDLSSMQ 493

Qy      248 DEDLVRLYNQELVERWNLAASDILRIKSMUKBGGVYLDVDIILPGIOPDLFKSK- 303
Db      494 DRTNTYNYEMEMLLRWNPYPAASQIRMYMLKEGIGIYTDLDMMVPOYSPOVLQMINBIGD 553

Qy      304 ----- 303
Db      554 RFEEUPIRAVSGVURLANGESGITIDNIAODISKUTRSDRTOLEKLLTDIENKOK 613

Qy      304 -----PDSIT-----NTSWEMIKLEAMTKYKEVTPGYTSKNFDM 337
Db      614 SSGTSKFSLQFMAPDSVRDFMFILORHKKWSICNVVRLNGLM-----MAHKGSA 665

Qy      338 LDEEV--QRSFESALSKSD--KSEIFLPDDIKVSPLEVKTA-----F 377
Db      666 VDAVIKQORQAVRELKSLRETVLSGSEFFKTLDDLKHLNHRKALICGHVLDFLKSLSFSE 725

Qy      378 ANNSVINQALISLSDVCSLDVINOIKRVYKILNDNLNPSINSGTGFNTMKLFSKDLAS 437
Db      726 RQDTIITAGALSTIGIT--GPDLIVDEMKTYFRSLGP--IGKDYLGSKGLGN--KAPLGSYNK 781

Qy      438 ISNEDNMVF--MIKITNVLKVGAPDVRSTINLSGPGVYTGAYQDLIMPDKDNTNHLLE 495
Db      782 IVKEGQTFDMLHPVT-----VG-ANDVTPADESTWCGIKORCLAEILL--FSDESK----FR 831

Qy      496 PELRNPEFPK-----TKISQTEOEITSLWSFNQARAKS-----QFEYKXGYFEGALGED 546
Db      832 PET-----PKGISRTKIDKXT--FTKLWSEQSKRALSPELLDRFNE-----LIEA 874

Qy      547 DNLDFAQNTVLDKD--YVSKKILSMKTRNEYIHYIVQ-----LQGDKISYEASCNLFSDK 601
Db      875 KNFDLIKFTDVRDALYVLRNQCDDASAAVFSLQLQLAELLRGTSFFPIANQLNVFPD- 933

Qy      602 PYSSILYQKNIEGSETAYVYVADAEIKDYRIPYQISKNKNIKLTFIGHCKSE---- 657
Db      934 -----LNRNIAS-----SLEKTIKLYLOSHSQCTETVIW 961

Qy      658 -----ENTDTFA-----NLDVDSLSEIETILNAXADISPKYIEI-----NLLGCNM 700
Db      962 HSALSDRVMLFRLMLATAERMRLVKESIFSKEETFLTTEBIKLNSWAEQLSKDNLGI-- 1019

Qy      701 FVSYSABEYYPGKLL-----LKIKORVSELMFISQDSITVSANQYEVRIINE--- 748
Db      1020 -----LSLEEV--DKLLDVTSSIMENSKLKRVSAL-----EDSIV--AGFYFLRELBETLS 1066

Qy      749 -----EGKREIL-----DHSGKWINKESIIKDISKYEIVSFNPKENK 786
Db      1067 EWRTRIPKDKRVIRMEGMDANQQPDERGRWYSQ-----LYDESFKKRVTPAVKIQ 1122

Qy      787 IIVKSKYHELSTLLOBIRNNANS-----SDIDLEKKYMLTECEINVASNI 832

```

Db 1123 ELAK-KYLNEQRVHVLDDYLTAKNPLFNRLHESGYAFSDLTETRYMLASYGIS----- 1176

Qy 833 DRQIVEGRIEEAKNLTSDSI-----NYIKNEFKLIESISDLYDLKHQNG---LDDSH 882

Db 1177 -----GIFSEGNLPSAELVNI IKTYVGGDYHMQDVLPKIYDMLASGAADLTNER 1230

Qy 883 FIFSED-----ISKTENGFRIRFINKETGNSIFITEKEIPS-----EY 921

Db 1231 FAAPESLRKNLEGLHGTDLTTPVDASVSGWGSF-GVENG-----VESDHTMISAPGF 1285

Qy 922 ATHISKEISNKTDFD-----NVNGKLIVKKNLDAAEVNTLNSAFI-----QSL 968

Db 1286 FNGASYGMQHSYALYSIHRHILGJSTSDIIRK-----ELESKGACGFVHEERFDSL 1338

Qy 969 IEYNTTKE--SLSNLSVAMKVQVY-AOLFSTGLNT-ITDASKVVE 1009

Db 1339 LKASSEKQYLSLTIEHKLSNQVHEAEAVSHLMMTALPGVCKIIE 1383

RESULT 14

JC6309

surface-located membrane protein lmp3 precursor - Mycoplasma hominis

C:Species: Mycoplasma hominis

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: JC6009

R:Ladefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.

J. Bacteriol. 178, 2775-2788, 1996

A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene system

A:Reference number: JC6009; MUID:96213016; PMID:8631664

A:Accession: JC6009

A:Molecule type: DNA

A:Residues: 1-1302 <LAD>

A:Cross-references: EMBL:X95601; NID:gl197335; PIDN:CAA64858.1; PID:gl197336

C:Genetics:

A:Gene: lmp3

A:Genetic code: SGC3

C:Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homolog.

C:Keywords: duplication; membrane protein

F:1-24/Domain: signal sequence #status Predicted <SIG>

F:25-1302/Product: surface-located membrane protein lmp3 #status predicted <MAT>

F:957-992/Domain: tetratricopeptide repeat homology <Ttl>

F:993-1026/Domain: tetratricopeptide repeat homology <TT2>

F:1089-1120/Domain: tetratricopeptide repeat homology <TT3>

F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match 6.1%; Score 315; DB 1; Length 1302;

Best Local Similarity 20.2%; Pred No. 0.00011;

Matches 235; Conservative 203; Mismatches 396; Indels 330; Gaps 53;

Qy 3 LVNKAQKQKMYVVR-FRIQDEYVAILNALAEYH-----NMSSSYVE--- 44

Db 169 LNKINAREELQSKIFNEKKQELXRLD-LDTKEVDFTKQKVFIETNINETSIEDIK 227

Qy 45 -KYLK-KDINNLTDNYTKYGRNKALKKFKYELTMEVLKKNLSLTPVEKNLHFTW 102

Db 228 NKIIEVEKATSSLSKILNT-----KQELQEF-ENIKKDLQDFINTKLNDAKYQ----- 276

Qy 103 IGGQINDTAINYQKVDNDVYKVFYDGNAFINTLTKTIVESAT-----NNTLESPEE 159

Db 277 ---SIKOKALDKINSLNGINKNSTIKETIKAGQNALIKAKEEAGLEKEKLDGQNIKDTLKE 333

Qy 160 NLNDEPFDYKFKWEIIVDKQHFIDYKQSIQENPEFIIDNI IKTYLSNEYSKDLE 219

Db 334 TINNAK-----EFKKLIDNDQKIVD--LKSNLUNE-SKADQ 368

Qy 220 ALNKYIEESLNKITAANGDIRNLEKFADEDLVRLYNQELVERWNLAASDILR-SMLKE 279

Db 369 SLSK-----DKESVESANDLLNKLIEYKILNFKNQEKAKFN-----ELEQTRK 414

Qy 280 DGGVYLDVDI-----LPGIQDLPFKSINKPDSITNTSWEMIKLEAIMKYKVIPTGTSKN 334

Db 415 NTFNFTDPEVKNPNVATLVKDLTNAKDKKSVTNSG---NKSDIIIAANEALIOALADAN 471


```

QY 776 EYISPNKEN-----KIIVSK-YLHELSTLLQEIRNNAN 809
Db : : : : :
836 LQWVSESONHSDINQCLNEVANIYNILKLNKIKKIIDKVKEYTSEIEKNKNINDELN 895
QY 810 SSDIDLEK---KWLTECEINVASNIDROIVEGRIE-----AKNLTSDSINYIKNEFK-- 860
Db : : : : :
896 NSEKVIKKIEGLSLKECRSKINSTLDDKDIDECIKNINVLKKNILNEBTN-ITNHFKNV 954
QY 861 -----LIESISDLYDLKHONGLDDSHFISFEDISK---TENGPRIRFIN 902
Db : : : : :
955 EYNNKIVLSNFNNIEMADNKSQYILEIKQNGTND-HDYNIKELKSHKXNGYK---T 1009
QY 903 KETGNSIFTETEKEIPSEYATHI-----SKEISNIKDTIFDNVNGKLVKKNLDAAH 954
Db : : : : :
1010 EADQNKKAIOKNKELFEQYKEVTVLLNKYVAVELKNKEDKT-KNDSKQIIEIK--DAH 1066
QY 955 EVNTLNSAFFIOSL-----IEYNTTKESLSNLSVAMKVQVYAQLFSTGLNITD 1003
Db : : : : :
1067 NYCTLESQKSEKQWNEIKNEKIHIEDEVANNDKSNKAIT-SIKVSEPEPKIINE 1123

```

Search completed! April 1, 2004, 16:45:02
Job time : 22.25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:44:46 ; Search time 34.125 Seconds
(without alignments)
7840.786 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1020
Perfect score: 5189
Sequence: 1 MNLVNRQAQLQKVVVFRIQ.....ITDASKVVELSPALDETID 1020

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1069545 seqs, 262320428 residues

Total number of hits satisfying chosen parameters: 1069545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4101	79.0	2366	14	US-10-011-366-10
2	4101	79.0	2366	15	US-10-354-774-10
3	4101	79.0	2366	15	US-10-271-012-10
4	2791	53.8	2710	14	US-10-011-366-6
5	2791	53.8	2710	15	US-10-354-774-6
6	2791	53.8	2710	15	US-10-271-012-6
7	2204	42.5	556	12	US-10-463-957-1
8	2193	42.3	556	12	US-10-463-957-11
9	2190	42.2	556	12	US-10-463-957-12
10	2051.5	39.5	522	12	US-10-463-957-9
11	1979	38.1	500	12	US-10-463-957-3
12	1734	33.4	420	12	US-10-463-957-5
13	769	14.8	170	12	US-10-463-957-7
14	422.5	8.1	3169	14	US-10-114-170-257
15	290.5	5.6	4688	12	US-10-282-122A-76865

16	279	5.4	1178	12	US-10-282-122A-52434	Sequence 52434, A
17	276.5	5.3	1182	12	US-10-282-122A-53445	Sequence 53445, A
18	275.5	5.3	1279	12	US-10-282-122A-52455	Sequence 52455, A
19	270	5.2	1196	12	US-10-282-122A-52737	Sequence 52737, A
20	265	5.1	4620	12	US-10-282-122A-68921	Sequence 68921, A
21	258	5.0	1090	15	US-10-369-493-18439	Sequence 18439, A
22	257.5	5.0	1162	15	US-10-452-024-113	Sequence 113, App
23	257.5	5.0	1162	15	US-10-452-024-114	Sequence 114, App
24	257.5	5.0	2184	14	US-10-304-095-6	Sequence 6, Appli
25	256	4.9	903	12	US-10-282-122A-5328	Sequence 5328, A
26	255	4.9	1875	15	US-10-369-493-22285	Sequence 22285, A
27	253.5	4.9	1184	12	US-10-282-122A-53254	Sequence 53254, A
28	252	4.9	1948	14	US-10-032-585-7611	Sequence 7611, Ap
29	249	4.8	1679	15	US-10-369-493-22080	Sequence 22080, A
30	245.5	4.8	1275	12	US-10-282-122A-53272	Sequence 53272, A
31	245.5	4.7	1103	12	US-10-282-122A-76866	Sequence 76866, A
32	245.5	4.7	1881	14	US-10-032-585-7646	Sequence 7646, Ap
33	245	4.7	1639	14	US-10-087-464-10	Sequence 10, Appl
34	244.5	4.7	6641	12	US-10-282-122A-70580	Sequence 70580, A
35	244	4.7	1143	15	US-10-369-493-11081	Sequence 11081, A
36	243	4.7	1928	15	US-10-369-493-22025	Sequence 22025, A
37	242	4.7	1005	15	US-10-369-493-1061	Sequence 1061, Ap
38	241	4.6	1494	12	US-10-282-122A-47128	Sequence 47128, A
39	239.5	4.6	1957	15	US-10-369-493-2070	Sequence 2070, Ap
40	239	4.6	1837	15	US-10-369-493-22734	Sequence 22734, A
41	237.5	4.6	1162	15	US-10-452-024-115	Sequence 115, App
42	237.5	4.6	5005	12	US-10-282-122A-76871	Sequence 76871, A
43	237	4.6	1163	12	US-10-282-122A-51864	Sequence 51864, A
44	236.5	4.6	1790	15	US-10-369-493-1586	Sequence 1586, Ap
45	235.5	4.5	1155	12	US-10-282-122A-53419	Sequence 53419, A

ALIGNMENTS

RESULT 1

US-10-011-366-10
; Sequence 10, Application US/10011366
; Publication No. US20030054493A1
; GENERAL INFORMATION:

APPLICANT: Williams, James A.

APPLICANT: Kink, John A.

TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES

OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE

DISEASE

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/011.366

FILING DATE: 16-NO. US20030054493A1-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/957.310

FILING DATE: 23-OCT-1997

APPLICATION NUMBER: US 08/329.154

FILING DATE: 24-OCT-1994

APPLICATION NUMBER: US 08/161.907

FILING DATE: 02-DEC-1993

APPLICATION NUMBER: US 07/985.321

FILING DATE: 04-DEC-1992

APPLICATION NUMBER: US 07/429.791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-011-366-10

Query Match 79.0%; Score 4101; DB 14; Length 2366;
Best Local Similarity 77.6%; Pred. No. 4.8e-250;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;

QY 1 MNLVNAQLOQKVVYKFRIOEDVEYVAILNALEVEYHNMSSESVVEKYKLKDKNNLTDNYL 60
DB 1 MSLVNRKQLEKMANVRFTQEDVEYVAILDALEVEYHNMSENTVVEKYKLKDKINSLTDIYI 60
QY 61 NTVYKSGRNKALKKFKKEYLTWVLELKNNSLTPVEKNLHFWIGQINDTAJYINQWKD 120
DB 61 DTYKSGRNKALKKFKKEYLTWVLELKNNSLTPVEKNLHFWIGQINDTAJYINQWKD 120
QY 121 VNSDVTYVFDVSNAPLNTLTKTIVESATNTNLTESFRENLDNDFYKFRKRMIIY 180
DB 121 VNSDVTYVFDVSNAPLNTLTKTIVESATNTNLTESFRENLDNDFYKFRKRMIIY 180
QY 181 DKQHFIDYKQSBENPEFIIDNIITYLSNEYSKOLEALNKYIEBSLNKITANNNDI 240
DB 181 DKQKFNINYKQARENFELIIDVITYLSNEYSKIDEINTYIEBSLNKITQNSGNDV 240
QY 241 RNLEKFAEDLVRLNQELVERWNLAAASDLIRLSMLKEDGGVLDVLDLPICQPDLFKS 300
DB 241 RNFEFKNGESFNLYEQELVERWNLAAASDLIRLSALKEIGMYLDVDMPLFGIQPDLFES 300
QY 301 INKPDISNTSEMIKLAIRKVEYIPGYTSKNFMDLDESVQSFSSALSSKSKSEIF 360
DB 301 IEKPSVTVDFWMTKLAIRKVEYIPGYTSKNFMDLDESVQSFSSALSSKSKSEIF 360
QY 361 LPLDIDKVPLEVKIAFANNVINQALISLKDSCSDIVINQIKRYKILNDNLNPSINE 420
DB 361 SSLGMEASPLEVKIAFANNVINQALISLKDSCSDIVINQIKRYKILNDNLNPSINE 420
QY 421 GDFNTWKIFSDKLASISNEDNMFMTKITNYLKVGPAPDVRSTINLSGPGVTGAYQD 480
DB 421 DNDFTNTNTFIDSIMAEANADNGRFMELGKYLRVGFPDVKTTINLSGPEAYAAAYQD 480
QY 481 LLMFKDNTNTHLPELNRNPPKTKISQTEQETSLNFPNOARAKSQFEYKKGYYE 540
DB 481 LLMFKEGSMNTHLEADLRNFEISKTNTISQTEQEMASLWFSDDARAKAQFEYKRYNFE 540
QY 541 GGLGDDNDLPQNVLDKVVYSKILGSMKTRKEYIHYIVQLQGDKISYEASCNLFPSK 600
DB 541 GGLGDDNDLPQNVLDKVVYSKILGSMKTRKEYIHYIVQLQGDKISYEASCNLFPSK 600
QY 601 DPYSSILYQKNTEGSETAYYYVADAEIKDKYRIPYQISNKNRIKLTFFGHGKSEFNT 660
DB 601 TPYDSVLFQKNTEGSETAYYYVADAEIKDKYRIPYQISNKNRIKLTFFGHGKSEFNT 660
QY 661 DTFANLVDVLSSEETILNLAKADISPKYIEINLLGCNMFYSISABEYTPGKLLIKIK 720
DB 661 DIFAGFDVDSLSTEAAIDAKEDISPKSIEINLLGCNMFYSINVEEYTPGKLLIKIK 720
QY 721 DRVSELMPSISODSITVSANQVEYVINEEGKEEIIIDHSGKMNKEESIIKDISKEYISF 780
DB 721 DKISELMPSISODSITVSANQVEYVINEEGKEEIIIDHSGKMNKEESIIKDISKEYISF 780
QY 781 NPKENKIIVKSKYLHELSTLLOEIRNANSDDILEKKVMLTECEINVASNIDQIVEGR 840

DB 781 NPKENKIIVKSKYLHELSTLLOEIRNANSDDILEKKVMLTECEINVASNIDQIVEGR 840
QY 841 IBEAKNLTSDSINYIKNEFKLIESDSLYDLKHQNG:DDSHFISFEDISKTENGFRIRF 900
DB 841 IBEAKNLTSDSINYIKNEFKLIESDSLYDLKHQNG:DDSHFISFEDISKTENGFRIRF 900
QY 901 INKETGNSIFIEETEKEIESEYATHISKEISNIKOTIFDNVNGKLVKYNLDAAEHVNTLN 960
DB 901 INKETGNSIFIEETEKEIESEYATHISKEISNIKOTIFDNVNGKLVKYNLDAAEHVNTLN 960
QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTTIDASKVVELVSTALDETID 1020
DB 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTTIDASKVVELVSTALDETID 1020

RESULT 2
US-10-354-774-10
Sequence 10, Application US/10354774
Publication No. US20030215468A1
GENERAL INFORMATION:
APPLICANT: Williams, James A.
Thalley, Bruce S.
TITLE OF INVENTION: Multivalent Vaccine For Clostridium Botulinum Neurotoxin
NUMBER OF SEQUENCES: 82
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/354,774
FILING DATE: 30-Jan-2003
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/704,159
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-02304
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-354-774-10

Query Match 79.0%; Score 4101; DB 15; Length 2366;
Best Local Similarity 77.6%; Pred. No. 4.8e-250;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;

QY 1 MNLVNAQLOQKVVYKFRIOEDVEYVAILNALEVEYHNMSSESVVEKYKLKDKNNLTDNYL 60
DB 1 MSLVNRKQLEKMANVRFTQEDVEYVAILDALEVEYHNMSENTVVEKYKLKDKINSLTDIYI 60
QY 61 NTVYKSGRNKALKKFKKEYLTWVLELKNNSLTPVEKNLHFWIGQINDTAJYINQWKD 120
DB 61 DTYKSGRNKALKKFKKEYLTWVLELKNNSLTPVEKNLHFWIGQINDTAJYINQWKD 120

QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESAANNLTLESFRENLDNPEFDYKFKRMEIYY 180
DB 121 VNSDYNVNVFYDSNAFLINTLTKTIVESAANNLTLESFRENLDNPEFDYKFKRMEIYY 180
QY 181 DKQKHFDYKSKQIEENPEFIIDNIKTYSLSNEYSKDEALNKYIESLNKINTANGNDI 240
DB 181 DKQKNFINYKQAEENPEFIIDNIKTYSLSNEYSKDEALNKYIESLNKINTANGNDI 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILIRISMLKEDGGVYLDVILPGIQDLPFKS 300
DB 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILIRISMLKEDGGVYLDVILPGIQDLPFKS 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEVOQSFEALSXSXSKSEIF 360
DB 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKNFMDLDEVOQSFEALSXSXSKSEIF 360
QY 361 LPLDDIKVSPLEVKIAFANNSVINQALISLKDYSQDLVINQIKRYKILNDNLNPSINE 420
DB 361 SSLGDMEASPLEVKIAFANNSVINQALISLKDYSQDLVINQIKRYKILNDNLNPSINE 420
QY 421 GTDFTNTMTKIFSDKLASISNEDMMFMKITYLVKGFAPDVRSINLSGPGVYTCAYOD 480
DB 421 DNDFTNTMTKIFSDKLASISNEDMMFMKITYLVKGFAPDVRSINLSGPGVYTCAYOD 480
QY 481 LLMFKDGNSTNHLLEPELRNPFPPKTKISQTEQITSLMSFNQARAKSOFEEYKKGYPE 540
DB 481 LLMFKDGNSTNHLLEPELRNPFPPKTKISQTEQITSLMSFNQARAKSOFEEYKKGYPE 540
QY 541 GALGEDDNLDPQNTVLDKQVSKILSMKTRKKEYIHYIVQLQDKISYEASCNLFK 600
DB 541 GSGLEDNLDPSQNVIVVKEYLEKISLARSSRGYIHYIVQLQDKISYEASCNLFK 600
QY 601 DPYSILYQKNIQESGTAHYIVVADAEIKIDKIRIPYQISNKNKILTIHGKGEFNT 660
DB 601 TPYDSVLQKNIQESGTAHYIVVADAEIKIDKIRIPYQISNKNKILTIHGKGEFNT 660
QY 661 DTEANLDVSSLSIEITNLAKADISPKYIEINLLGNMFSYSISAEETYPCKLLKIK 720
DB 661 DIFAGFDVDSLUSTIEAIDAIDAKEDISPKYIEINLLGNMFSYSISAEETYPCKLLKIK 720
QY 721 DRVSELMPSISQDSITVSANQYVINEEGKREILDHSGKWINKEESIIKDISKEYISF 780
DB 721 DKISELMPSISQDSITVSANQYVINEEGKREILDHSGKWINKEESIIKDISKEYISF 780
QY 781 NPKENKIIVKSKYLHEISTLLOEIRNNANSDDIDLEKVMMLTECEINVASINDRQIVEGR 840
DB 781 NPKENKIIVKSKYLHEISTLLOEIRNNANSDDIDLEKVMMLTECEINVASINDRQIVEGR 840
QY 841 IEEAKNLTSDSINVIKNEFKLIESISDLYDLKHONGLDSDHSPFISFEDISKTENGPRIF 900
DB 841 IEEAKNLTSDSINVIKNEFKLIESISDLYDLKHONGLDSDHSPFISFEDISKTENGPRIF 900
QY 901 INKGTGNSIFTEKEIFSEVATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHEVNTLN 960
DB 901 INKGTGNSIFTEKEIFSEVATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHEVNTLN 960
QY 961 SAFTIQSLIEYNTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAFIIQSLIEYNTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 3
US-10-271-012-10
; Sequence 10, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlien & Carroll

STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/271,012
FILING DATE: 15-Oct-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/704,159
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-02304
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-271-012-10
Query Match 79.0%; Score 4101; DB 15; Length 2366;
Best Local Similarity 77.6%; Pred. No. 4.8e-250;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;
QY 1 MNLVANKAQLQKVVYKRIQDEYVAIILNALLEEYHNMSESSWEKYKLKXINLNTDYL 60
DB 1 MSLVNRKQLERKAVNFRFQDEYVAIILNALLEEYHNMSESSWEKYKLKXINLNTDYL 60
QY 61 NTKKSGRKNALKKPKKEYLTMEVLKKNLSLTPVEKNLHFIWIGQINDTAINYNQWD 120
DB 61 DTKKSGRKNALKKPKKEYLTMEVLKKNLSLTPVEKNLHFIWIGQINDTAINYNQWD 120
QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESAANNLTLESFRENLDNPEFDYKFKRMEIYY 180
DB 121 VNSDYNVNVFYDSNAFLINTLTKTIVESAANNLTLESFRENLDNPEFDYKFKRMEIYY 180
QY 181 DKQKHFDYKSKQIEENPEFIIDNIKTYSLSNEYSKDEALNKYIESLNKINTANGNDI 240
DB 181 DKQKNFINYKQAEENPEFIIDNIKTYSLSNEYSKDEALNKYIESLNKINTANGNDI 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILIRISMLKEDGGVYLDVILPGIQDLPFKS 300
DB 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILIRISMLKEDGGVYLDVILPGIQDLPFKS 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEVOQSFEALSXSXSKSEIF 360
DB 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKNFMDLDEVOQSFEALSXSXSKSEIF 360
QY 361 LPLDDIKVSPLEVKIAFANNSVINQALISLKDYSQDLVINQIKRYKILNDNLNPSINE 420
DB 361 SSLGDMEASPLEVKIAFANNSVINQALISLKDYSQDLVINQIKRYKILNDNLNPSINE 420
QY 421 GTDFTNTMTKIFSDKLASISNEDMMFMKITYLVKGFAPDVRSINLSGPGVYTCAYOD 480
DB 421 DNDFTNTMTKIFSDKLASISNEDMMFMKITYLVKGFAPDVRSINLSGPGVYTCAYOD 480
QY 481 LLMFKDGNSTNHLLEPELRNPFPPKTKISQTEQITSLMSFNQARAKSOFEEYKKGYPE 540
DB 481 LLMFKDGNSTNHLLEPELRNPFPPKTKISQTEQITSLMSFNQARAKSOFEEYKKGYPE 540

541 GALGEDNDLPAQNTVLDKDVSVKILSSMKTRNKEVIHYIVQLQGDKISYEASCNLFK 600
 541 GSLGEDNDLPSQIVVDKEYLLEKISLARSSRGYIHYIVQLQGDKISYEACNLFAK 600
 601 DPYSILYOKNIESEYATYVYVADAETDKYRIPIYQISNKENIKLTIGHGKSEFNT 660
 601 TPYDSVLPQKNIEDSEIATYVYVADAETDKYRIPIYQISNKENIKLTIGHGKSEFNT 660
 661 DTFANLDVSLSSIEITILNAKADISPKYIEINLLGNMFSYSISAEETYPGKLLKIK 720
 661 DIFAGFDVSLSTIEAIDAIDAKEDISPKYIEINLLGNMFSYSINVEETYPGKLLKVK 720
 721 DRVSELMPSISQDSITVSAQYEVRIINEEGKREILDHSGKWINKESIIKDISKEYISF 780
 721 DKISELMPSISQDSIIVSAQYEVRIINEEGKREILDHSGKWINKESIIKDISKEYISF 780
 781 NPKENKIVKSKYLHELSTILOETRNANSSDIDLEKKWMLTECEINVASNIDQIVEGR 840
 781 NPKENKIVKSKNLPSTLLQEIARNNSSDIDLEKKWMLTECEINVASNIDQIVEER 840
 841 TEEAKNLSDSINYIKNPKFLESISDLYDKHQGLDSDHFTSFEDISKTENGPRIIF 900
 841 TEEAKNLSDSINYIKNPKFLESISDLYDKHQGLDSDHFTSFEDISKTENGPRIIF 900
 901 INKETSIGIFETEKIEPSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAHEVNTLN 960
 901 INKETSIGIFETEKIEPSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAHEVNTLN 960
 961 SAFFQSIETIYNTTKESLSNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDSTID 1020
 961 AAFQSIETIYNTTKESLSNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDSTID 1020

RESULT 4

US-10-011-366-6
 Sequence 6, Application US/10011366
 Publication No. US20030054493A1
 GENERAL INFORMATION:
 APPLICANT: Williams, James A.
 Kink, John A.
 TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
 OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
 DISEASE
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Medien & Carroll
 STREET: 220 Montgomery Street, Suite 2200
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/011,366
 FILING DATE: 16-NOV-2003
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/957,310
 FILING DATE: 23-OCT-1997
 APPLICATION NUMBER: US 08/329,154
 FILING DATE: 24-OCT-1994
 APPLICATION NUMBER: US 08/161,907
 FILING DATE: 02-DEC-1993
 APPLICATION NUMBER: US 07/985,321
 FILING DATE: 04-DEC-1992
 APPLICATION NUMBER: US 07/429,791
 FILING DATE: 31-OCT-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPD-01121
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 US-10-011-366-6

Query Match 53.8%; Score 2791; DB 14; Length 2710;
 Best Local Similarity 52.1%; Pred. No. 3.6e-167;
 Matches 534; Conservative 221; Mismatches 263; Indels 8; Gaps 5;

QY 1 MNLVNAQLQKQVYKFIQDEYVAIALNEEYNNMSESSVVEKYLKLDINNLTNLYL 60
 DB 1 MSLISKEELIKLAY-SIRPRENEYKTLTNDYNNKLTNNENKYLQKLINESIDVFM 59
 QY 61 NNYKSGRNKALKKPEVLTWEVLELKNLSLTPVEKNLHFTWIGQINDTAINYNQKD 120
 DB 60 NKYKTSNRVRLSNLKKOILKEVILIKNSNTSPVEKNLHFTWIGVSDIALEYIKQAD 119
 QY 121 VNSDVTWKVYDSNAFLNTLTKKTIVESATNNLTSPRENLMDFDYNKFKYRMEIY 180
 DB 120 INAEYNKLVYSEAFVNTLKAIVESTTEALQLLEEIQNFQDNKFKYKKEMEIY 179
 QY 181 DKQKFIIDYKQIEENPEFIIDNTIKYLSNEYKDLKALNKYTEESINKITANGNDI 240
 DB 180 DRQKRPINYSQINKPTVTIDDIKSHLVSEYNRDETLESYRTNSLRKINSNHGDI 239
 QY 241 RNLEKPADLVRNLNOELVERWNLAASDILIRISMLKEDGVYLDVLDLPGIQDLRKS 300
 DB 240 RANSLFTEQELLNYSQELLNGLAASDILIRLALKNFGVYLDVLDLPGIHSDLRKT 299
 QY 301 INKPSDITNTSWEMIKLEAIMKYKEVIPOYTSKNDMLDEEVORFEGALSSEKSKSEIF 360
 DB 300 ISRPSSIGLDREWMKLEAIMKYKINNYISENFDKLDQKLKDNFKLIESSEKSEIF 359
 QY 361 LPDDITKVPFLEVYKAFANNVINQALISLSDSYCDLVINQIKNRYKILNDNLNPSNE 420
 DB 360 SKLENLVSDLEIKAFALGVSINQALISLSDSYCDLVINQIKNRYKILNDNLNPAIES 419
 QY 421 GTDFNTMKIFSDKLASISNEDNMFMKITNYLVKGFAPDVRSSTINISGPPVYTGAD 480
 DB 420 DNNFTDTTKIFHDSLFNSATAENSMPLTKIAPYLVQGFMPPEARSTISLGGPAYSAYD 479
 QY 481 LLMFKDNSTNIHLLEPELNPEPKTKISQTEQITSLWSFNQARAKSQREYKGYFE 540
 DB 480 FINLQENTIEKTKASDLIEFKPEPENNLSQTEQITSLWSFNQARAKSQREYKGYFE 539
 QY 541 GALGEDNDLPAQNTVLDKDY-VSKKILSS--MKTRNKEYIHYIVQLQGDKISYEASCNL 597
 DB 540 GSLSEDNGVDNFKNATLDKYNLLNKKIPSNVVEAGSKYVYIYIQLQDDDI-SYEATCNL 599
 QY 598 FSKDPIYSSIIYQKNIIEGSETATYVYVAD--AEIKEDIKYRIPIYQISNKENIKLTIGHGK 655
 DB 600 FSKNPKNSIIQENM--NESAKSYFLSDDGSEILELNKYRIPELKNKPKVYVTFIGHGK 657
 QY 656 SEFNTTTFANLDVSLSSIEITILNAKADISPKYIEINLLGNMFSYSISAEETYPGKL 715
 DB 658 DEFNTSEFARLSVDSLNSISGFLDTIKLIDISPKVVEVNLNLCNNMFSYDFNVEETYPGKL 717
 QY 716 LKIKDRVSELMPISIQDSITVSAQYEVRIINEEGKREILDHSGKWINKESIIKDISK 775
 DB 718 LLSIMDKITSTLPVKNKSIITIGANOEYRINSEGRKELLASHGKWINKESIIKDISK 777
 QY 776 EYISFNPKNENKIVKSKYLHELSTILOETRNANSSDIDLEKKWMLTECEINVASNIDRQ 835
 DB 778 EYIFFDSIDNKLKAKSKNIPGLASISEDIKTLILDASVSPDKFILNKLNISSIGDY 837

Thalley, Bruce S.
TITLE OF INVENTION: Multivalent Vaccine For Clostridium Botulinum Neurotoxin
NUMBER OF SEQUENCES: 82
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/271.012
FILING DATE: 15-Oct-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/704.159
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-02304
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-271-012-6

Query Match 53.8%; Score 2791; DB 15; Length 2710;
Best Local Similarity 52.1%; Pred. No. 3.6e-167;
Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;
QY 1 MNLVNAQLOKQVYKFIQDEVAIINLAEYHNMSSESVKYLKLDINLTDNYL 60
DB 1 MSLISKEELIKLAY-SIRPRENEYKTLITLDNEYKLTNNENKYLQKLKLNESIDVFM 59
QY 61 NTKYSGENKALKKFEVLTVELEKNSLTPVEKHLFIWIGQINDTAINYQWAD 120
DB 60 NKYTSRNRALSNLKDKILKEVILIKNSNTPVEKNLHFVWIGVEVSDIALEVIKOWAD 119
QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNNTLESFRENLDPEFDYKFKRMEIY 180
DB 120 INAEYIKWDSFAFLVNTLKAIVESTTEALQLLEEIQNFQDNMKEYKGEFIY 179
QY 181 DKQHFIDYKYSQIENPEFIIDNIITKYLNEYSKOLEALNKYIBESLNKITANNNDI 240
DB 180 DRQKFEINYKSIQKPTVPTDIDIKHGLVSEYNRDETVELSYRTNSLRKINSNHGIDI 239
QY 241 RNLEKPAEDLVRLNQELVERAWNLAASDILRISMLKEDGGVLDVLDILGCIQDPLFKS 300
DB 240 RANSLFTEQELNITYSQELNARNGNLAAASDILRLALKNFGVYLDVDMLEFGIHSDFKT 299
QY 301 INKPDSTNTSWEMIKLEAIMKYKVIPIGYTSKNDMLDEEVQSFESALSKDKKSEIF 360
DB 300 ISRPSIGLDRWEMIKLEAIMKYKVIINNYTSFNFDKLDQDKONFKLIIESKESEIF 359
QY 361 LPLDDIKVSPLEVKIAFANNVINQALISLKDYSCLVINGIKRYKILNDNLNPSINE 420
DB 360 SKLENLNSDLKIKAFALGVSINQALISKQSYLTNLVIEQVKRYQFLNCHLNPALES 419
QY 421 GTDFNTMTKIFSDKXIASINENNMFMKIKITNYLVKGFAPDVRSTINISGPGVYTGAVQD 480

RESULT 7
US-10-463-957-1
; Sequence 1, Application US/10463957
; Publication NO. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyres, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-1

Query Match 42.5%; Score 2204; DB 12; Length 556;
Best Local Similarity 75.0%; Pred. No. 5e-131;
Matches 417; Conservative 66; Mismatches 73; Indels 0; Gaps 0;
QY 1 MNLVNAQLOKQVYKFIQDEVAIINLAEYHNMSSESVKYLKLDINLTDNYL 60
DB 1 MSLVNRKQLEKMANVRFTQDEBYAIDALBEYHNMSSESVKYLKLDINLTDIYI 60
QY 61 NTKYSGRNLKALKKFEVLTVELEKNSLTPVEKNLHFVWIGQINDTAINYQWAD 120

```

Db 61 DTYKSGRNKALKKFKFYLVTVELEKNNLTPVEKNLHFVWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVPYDSNAFLINTLTKTIVESATNTNTLESFRENLDNPFDPYKFKRMEIYY 180
Db 121 VNSDYVNVFYDSNAFLINTLTKTIVESAINDTLESFRENLDNPFDPYKFKRMEIYY 180
QY 181 DKQKHFDYKSOIENPEFIIDNIKTYSNEYSKOLEALNKYIEESLNKITANGNDI 240
Db 181 DKQKNFYNYKQREENPELIIDDIKVTYLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240
QY 241 RNLEKFADELDVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFKNGESFNLYEQELVERWNLAASDILRISALKETGGMYLDVDMPLPGIQDPLFES 300
QY 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEDEEVORSFESALSSKSKSEIF 360
Db 301 IEKPSSTVDFWEMTKLEAIMKYKEYIPEYTSHEFDMLEDEEVORSFESALSSKSKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINQKRYKILNDMLNPSINE 420
Db 361 SSLGDMASPLEVKIAFANNVINOALISLSDSYCSNLIVKQIENRYKILNNSLNPASE 420
QY 421 GTDFNTMTKIFSDKLASINEDNMFMKITYNLYKGFAPDVRSTINLSGPGVYTCAYOD 480
Db 421 DNDFTNTTNTFIDSIMAEANADNGRFMMELGKYLVRGFPDVKTTINLSGPEYAAAYOD 480
QY 481 LLMFKDNSTNIHLEPELRNFEPPKTKISQLTEQETISLWSFNQARAKSQFEEYKKGYPE 540
Db 481 LLMFKEGSNVNIHLEADLRNFEISKTNISQSTQEMASLWSFDDARAKAQFEEYKKNYFE 540
QY 541 GALGEDNDLDPQNTV 556
Db 541 GSLGEDNDLDFSQNV 556

```

```

RESULT 8
US-10-463-957-11
; Sequence 11, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spvres, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Completely synthesized
; NAME/KEY: MISC FEATURE
; LOCATION: (395)..(395)
; OTHER INFORMATION: Xaa at position 395 is ala, asp, glu, phe, gly, his, ile, lys,
; OTHER INFORMATION: leu, met, asn, pro, gln, arg, ser, thr, val, trp, or tyr.
US-10-463-957-11

```

```

Query Match 42.3%; Score 2193; DB 12; Length 556;
Best Local Similarity 74.8%; Pred. No. 2.5e-130;
Matches 416; Conservative 66; Mismatches 74; Indels 0; Gaps 0;

QY 1 MNLVNAQLOKQVYVYKFIQDEYVAILNALLEEYHNMSSESVVEKYLKLDINLNTDYL 60
Db 1 MSLVNRKQLEKQVNVFRFTQDEYVAILDALEEYHNMSSENTVVEKYLKLDINSLTDIYI 60
QY 61 NTKYKSGRNKALKKFKFYLVTVELEKNNLTPVEKNLHFVWIGQINDTAINYNQWKD 120

```

```

Db 61 DTYKSGRNKALKKFKFYLVTVELEKNNLTPVEKNLHFVWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVPYDSNAFLINTLTKTIVESATNTNTLESFRENLDNPFDPYKFKRMEIYY 180
Db 121 VNSDYVNVFYDSNAFLINTLTKTIVESAINDTLESFRENLDNPFDPYKFKRMEIYY 180
QY 181 DKQKHFDYKSOIENPEFIIDNIKTYSNEYSKOLEALNKYIEESLNKITANGNDI 240
Db 181 DKQKNFYNYKQREENPELIIDDIKVTYLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240
QY 241 RNLEKFADELDVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFKNGESFNLYEQELVERWNLAASDILRISALKETGGMYLDVDMPLPGIQDPLFES 300
QY 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEDEEVORSFESALSSKSKSEIF 360
Db 301 IEKPSSTVDFWEMTKLEAIMKYKEYIPEYTSHEFDMLEDEEVORSFESALSSKSKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINQKRYKILNDMLNPSINE 420
Db 361 SSLGDMASPLEVKIAFANNVINOALISLSDSYCSNLIVKQIENRYKILNNSLNPASE 420
QY 421 GTDFNTMTKIFSDKLASINEDNMFMKITYNLYKGFAPDVRSTINLSGPGVYTCAYOD 480
Db 421 DNDFTNTTNTFIDSIMAEANADNGRFMMELGKYLVRGFPDVKTTINLSGPEYAAAYOD 480
QY 481 LLMFKDNSTNIHLEPELRNFEPPKTKISQLTEQETISLWSFNQARAKSQFEEYKKGYPE 540
Db 481 LLMFKEGSNVNIHLEADLRNFEISKTNISQSTQEMASLWSFDDARAKAQFEEYKKNYFE 540
QY 541 GALGEDNDLDPQNTV 556
Db 541 GSLGEDNDLDFSQNV 556

```

```

RESULT 9
US-10-463-957-12
; Sequence 12, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spvres, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 12
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-12

```

```

Query Match 42.2%; Score 2190; DB 12; Length 556;
Best Local Similarity 74.8%; Pred. No. 3.8e-130;
Matches 416; Conservative 66; Mismatches 74; Indels 0; Gaps 0;

QY 1 MNLVNAQLOKQVYVYKFIQDEYVAILNALLEEYHNMSSESVVEKYLKLDINLNTDYL 60
Db 1 MSLVNRKQLEKQVNVFRFTQDEYVAILDALEEYHNMSSENTVVEKYLKLDINSLTDIYI 60
QY 61 NTKYKSGRNKALKKFKFYLVTVELEKNNLTPVEKNLHFVWIGQINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKFKFYLVTVELEKNNLTPVEKNLHFVWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVPYDSNAFLINTLTKTIVESATNTNTLESFRENLDNPFDPYKFKRMEIYY 180
Db 121 VNSDYVNVFYDSNAFLINTLTKTIVESAINDTLESFRENLDNPFDPYKFKRMEIYY 180

```

QY 181 DKQHFIDYKSOEENPEFIIDNIKTYSNEYSKDLALNKYIBESLNKITANNNGDI 240
 Db 181 DKQNFINYKAQREENPELIIDIVKTYLSNEYSKEIDELNTYIBESLNKITQNSGNDV 240
 QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIOPDLFKS 300
 Db 241 RNFBEKNGESFNLYEOELVERWNLAAASDILRISALKEIGGMVYLDVMDLPGIOPDLFES 300
 QY 301 INKPDSTINTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSSKSDKSEIF 360
 Db 301 IEKPSSTVTDVFWEMTKLEAIMKYKEYIPGYTSEHFDLDEEVQSFESVLAASKDSEIF 360
 QY 361 LPLDDIKVSPLEVKIAFANNVINQALISLSDKSCYSDLVINQIKRYKILNDNLNPSINE 420
 Db 361 SSLGMEASPLEVKIAFNKGIINOGLISVXDSYCSNLIVKQIENRYKILNNSLNPAISE 420
 QY 421 GTFDNTWKIPSDKLASISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTGAYQD 480
 Db 421 DNDFTNTTFTDSTMAEANADNGRFMMELGKYLVRGFPDVKITINLSGPEAYAAAYQD 480
 QY 481 LLMFKDNSTNHLLEPELRNPEFPKTKISOLTEQELTSLWSFNQARAKSQEYKKGYFE 540
 Db 481 LLMFKEGSNVTHLLEADLRNFEISKTNISQTEQEMASLWSFDDARAKAQEYKRYNFFE 540
 QY 541 GALGEDDNLDPQNTV 556
 Db 541 GSLGEDDNLDFSQNV 556

RESULT 10
 US-10-463-957-9
 ; Sequence 9, Application US/10463957
 ; Publication No. US20040028705A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ballard, Jimmy D.
 ; APPLICANT: Spyres, Lea M.
 ; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
 ; FILE REFERENCE: 5820.637
 ; CURRENT APPLICATION NUMBER: US/10/463,957
 ; CURRENT FILING DATE: 2003-06-17
 ; PRIOR APPLICATION NUMBER: 60/389,685
 ; PRIOR FILING DATE: 2002-06-17
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 9
 ; LENGTH: 522
 ; TYPE: PRT
 ; ORGANISM: Clostridium difficile
 US-10-463-957-9

Query Match 39.5%; Score 2051.5; DB 12; Length 522;
 Best Local Similarity 73.8%; Pred. No. 2e-121;
 Matches 385; Conservative 66; Mismatches 70; Indels 1; Gaps 1;

QY 35 HNMSSSVVEKYLKLDINLNDYNTYKSGRNKALKKFKYELTMEVLELKNLSLTPV 94
 Db 1 HNMSENTVVEKYLKLDINSLDITDYIDTYKSGRNKALKKFKYELVIBELEENSLTPV 60
 QY 95 EKNLHFIWIGQNDTALNYINQKDVNSDYTVKVFYDSNAFLINTLTKKTIVESATNNTL 154
 Db 61 EKNLHFIWIGQNDTALNYINQKDVNSDYTVKVFYDSNAFLINTLTKKTIVESASNDTL 120
 QY 155 ESFRENLDPEFDYKFKRMEIYDKQHFIDYKSOEENPEFIIDNIKTYSNEY 214
 Db 121 ESFRENLDPEFNHTAFFRKMGIIDYKQNFINYKKAQKEENPDLIIDIVKTYLSNEY 180
 QY 215 SKOLEALNKYIBESLNKITANNNGDI RNLEKFADEDLVRLYNQELVERWNLAAASDILRI 274
 Db 181 SKOIDEALNAYIBESLNKVTNSGNDVRNPEEFTGTVFNLYEOELVERWNLACASDILRV 240
 QY 275 SMLKEDGGVYLDVILPGIOPDLFKSINKPDSI-TNTSWEMIKLEAIMKYKEYIPGYTSK 333
 Db 241 AILKNIGGVYLDVMDLPGIHPDLFKINKPDSVKTAVDWEEMQLEAIMKHKEYIPEYTSK 300

QY 334 NFDMLDEEVQSFESALSSKSDKSEIFLPLDDIKVSPLEVKIAFANNVINQALISLSDK 393
 Db 301 HFDTLDEEVQSFESVLAASKDSEIFLPLGDIIVSPLEVKIAFAGSIIINQALISAKDS 360
 QY 394 YCSDLVINQIKRYKILNDNLNPSINEGTDFNTWKIPSDKLASISNEDNMFMKITNY 453
 Db 361 YCSDLLIKQIQRYKILNDTLGPIISQNDNTNTWNGESLGALEENISFAKIGSY 420
 QY 454 LKVGFPAPDVRSTINLSGPGVYTGAYQD LLMFKDNSTNHLLEPELRNPEFPKTKISOLTE 513
 Db 421 LRVGFPEANTVTITLSGPIYAGAYKDLITFKEMSIDTISLSSELNPEFPKVNISQATE 480
 QY 514 QBITSLWSFNQARAKSQEYKKGIFEGALGEDDNLDPQNT 555
 Db 481 QEKNSLWQNEBRAKIQEYKKNYFEGALGEDDNLDFSQNT 522

RESULT 11
 US-10-463-957-3
 ; Sequence 3, Application US/10463957
 ; Publication No. US20040028705A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ballard, Jimmy D.
 ; APPLICANT: Spyres, Lea M.
 ; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
 ; FILE REFERENCE: 5820.637
 ; CURRENT APPLICATION NUMBER: US/10/463,957
 ; CURRENT FILING DATE: 2003-06-17
 ; PRIOR APPLICATION NUMBER: 60/389,685
 ; PRIOR FILING DATE: 2002-06-17
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 500
 ; TYPE: PRT
 ; ORGANISM: Clostridium difficile
 US-10-463-957-3

Query Match 38.1%; Score 1979; DB 12; Length 500;
 Best Local Similarity 75.0%; Pred. No. 7.2e-117;
 Matches 375; Conservative 60; Mismatches 65; Indels 0; Gaps 0;

QY 1 MSLVNRKALQKRVYKFRIOEDYVAILNALBEEYHNMSSSVVEKYLKLDINLNDY 60
 Db 1 MSLVNRKQLEKMANVFRIOEDYVAILDALBEEYHNMSENTVVEKYLKLDINSLDIYI 60
 QY 61 NTYKSGRNKALKKFKYELTMEVLELKNLSLTPVEKNLHFIWIGQINDTAINYINQWD 120
 Db 61 DTYKSGRNKALKKFKYELVTEVLELKNLTPVEKNLHFIWIGQINDTAINYINQWD 120
 QY 121 VNSDYTVKVFYDSNAFLINTLTKKTIVESATNNTLESFRENLDPEFDYKFKRMEIY 180
 Db 121 VNSDYNVNFYDSNAFLINTLTKKTIVESAINDTLESFRENLDPRFDYKFKRMEIY 180
 QY 181 DKQHFIDYKSOEENPEFIIDNIKTYSNEYSKDLALNKYIBESLNKITANNNGDI 240
 Db 181 DKQNFINYKAQREENPELIIDIVKTYLSNEYSKEIDELNTYIBESLNKITQNSGNDV 240
 QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIOPDLFKS 300
 Db 241 RNFBEKNGESFNLYEOELVERWNLAAASDILRISALKEIGGMVYLDVMDLPGIOPDLFES 300
 QY 301 INKPDSTINTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSSKSDKSEIF 360
 Db 301 IEKPSSTVTDVFWEMTKLEAIMKYKEYIPEYTSSEHFDLDEEVQSFESVLAASKDSEIF 360
 QY 361 LPLDDIKVSPLEVKIAFANNVINQALISLSDKSCYSDLVINQIKRYKILNDNLNPSINE 420
 Db 361 SSLGMEASPLEVKIAFNKGIINOGLISVXDSYCSNLIVKQIENRYKILNNSLNPAISE 420
 QY 421 GTDFTNTWKIPSDKLASISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTGAYQD 480


```
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMBELGKLYRVGFPPDVKTITNLSGPEAYAAAYQD 480
Qy 481 LLMFKDNSTNIHLLEPELAN 500
Db 481 LLMFKEGSMNIHLLEADLRN 500

RESULT 12
US-10-463-957-5
; Sequence 5, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Spvires, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; PRIOR FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-5

Query Match 33.4%; Score 1734; DB 12; Length 420;
Best Local Similarity 78.3%; Pred. No. 1.7e-101;
Matches 329; Conservative 48; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MNLVNAQLOKQVYVFRIOEDYVAIINALEEYHNMSESSVVEKYKDKDINNLTNYL 60
Db 1 MSLVNRKQLEKMANVRFTQEDYVAILDALAEYHNMSESSVVEKYKDKDINSLTDIYI 60
Qy 61 NTYKSGRNKALKKFKLEYLTMEVLELKNNSLTPEVKNLHFIWIGGOINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKFKLEYLTMEVLELKNNSLTPEVKNLHFIWIGGOINDTAINYNQWKD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTLSEFRENLDNDFPDYKFKRMEIY 180
Db 121 VNSDYNVNFYDSNAFLINTLKTIVESAINDTLESFRENLDNDFPDYKFKRMEIY 180
Qy 181 DKQKHFDIYKSOIEENPEFIIDNIITKTVLSNEYSKDLALANKYIEESLNKITANGNDI 240
Db 181 DKQKNFNYKQAEENPELIIDIVKTVLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240
Qy 241 RNLEKFAEDLVLYNQELVERWNLAASDILRISMLKSGGVYLDVILPGIQDLPFS 300
Db 241 RNFEFPKNGSFNLYEQELVERWNLAASDILRISALKSIGGYLDVDMPLPGIQDLPFS 300
Qy 301 INKPDSTNTSWEMIKLEAIKMYKEVPGYTSKNFMDLDEEVQSFESALSKSDKSEIF 360
Db 301 IEPSSVTVDVFWMTKLEAIKMYKEVPEYTSFHFMDLDEEVQSFESVLSKSDKSEIF 360
Qy 361 LPLDDIKVPLEVKIAFANNSVNOALISLSDYSCLVINOIKNRYKILNDNLNPSINE 420
Db 361 SSLGDMEASPLEVKIAFNKSGIINQGLISVDSYCSNLIVKQIENRYKILNLSNLPASE 420

RESULT 13
US-10-463-957-7
; Sequence 7, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Spvires, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
```

```
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-7

Query Match 14.8%; Score 769; DB 12; Length 170;
Best Local Similarity 85.3%; Pred. No. 4.7e-41;
Matches 145; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MNLVNAQLOKQVYVFRIOEDYVAIINALEEYHNMSESSVVEKYKDKDINNLTNYL 60
Db 1 MSLVNRKQLEKMANVRFTQEDYVAILDALAEYHNMSESSVVEKYKDKDINSLTDIYI 60
Qy 61 NTYKSGRNKALKKFKLEYLTMEVLELKNNSLTPEVKNLHFIWIGGOINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKFKLEYLTMEVLELKNNSLTPEVKNLHFIWIGGOINDTAINYNQWKD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTLSEFRENLDNDFPDYKFKRMEIY 170
Db 121 VNSDYNVNFYDSNAFLINTLKTIVESAINDTLESFRENLDNDFPDYKFKRMEIY 170

RESULT 14
US-10-114-170-257
; Sequence 257, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 257:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3169 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

SEQUENCE DESCRIPTION: SEQ ID NO: 257;		1043	KPEFVKSE-ISSAKSIFDDIKNK-YITDPETKENVLYHQDSDIKERIAFLDISHVAYP	1100
US-10-114-170-257		839	GRIEAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHQNGL	878
Query Match		1101	GSLLKQLS---GYFSDNIIAEVLLASGVSGHVSHGWYPAPSDKLELLRHTK	1156
Best Local Similarity		879	DDSHISPEDISKTEGFRIFINKETGNSI---FIETEKEIFSEYATHISKEISNIKOT	935
Matches 258; Conservative 197; Mismatches 450; Indels 323; Gaps 54;		1157	SNSEWI---EKITP-----YVDILSDNVSNVLRPFLSEQKILANDIKLEISKSVS---EQ	1207
QY	14	YVKFERI---QDEYVAILNAL---EEYHNMSESSVVEKYKLKLDNLTNYLNTYKKSGR	68	
DB	160	YIKIRKRGABDQTTTOSLIINELLGVDRNTI-PFKISELNDIIHSEYENQIKNSR	218	
QY	69	N--KALKKFEYLWLEVLKNS-----LTPVKNLHFIMIGQ	106	
DB	219	KGIELYKQGLSSLLNDKNGKQLSDNASKIINLLGIEVQSHKVDIEPFIHAWVAGA	278	
QY	107	INDRAININQKVNDSYTVKVFDSNAP-----LINTLKKT-----	144	
DB	279	PPDNTFSYITAFNTYKDYTYLLMTDPNFAAKFGILKNIAMYAIMLRNTNPLAE	338	
QY	145	-----IVESATNNTLE--SFRENLDPEFY-----NKPYRKRMEIIVDKQKHF	187	
DB	339	EMNEVILKIQNIQNETIEFKTRERLKELENRYKSLTSEYKTEKENVFFLESIMGMQDNYF	398	
QY	188	DYKSOIEENPEFI-----IDNIUK--TYLSNEYSKOLEALNKYIEESLNKITANNGN--	238	
DB	399	TYCISNGISNTDDISRLDFLWVTKLSPVQNDFKSVTEKNKRDIDLLKNTISQKGDGF	458	
QY	239	---DIRNLEKFADELDVLYNQELVERWNLAASDIIRISMLKEDGGVYLDVDILPGIQP	295	
DB	459	QLRDINTLESFKPKQDYFFYQEMLLRWYAAASQVRINILKEYGGIYTDITDILPAYSD	518	
QY	296	DLPKINPKDSTINTSWEMIKLEAMKYK--EYIPG--YTSKNFMDLDEEVQSFESALS	351	
DB	519	KVSQIINE-KSDKRFEDLRLRIISELSILKGEKYSIKH-DGUDETTLNQLNML-	575	
QY	352	SKSDSEIFLPUDDIKVSPLEVKIAFANNSVI-----NQALISLKDASYC	395	
DB	576	SEIEK-----LTDDY-FRPVETKVVRDFKIFKRYOKWTENTWIRGNFMFLTKGSKC	630	
QY	396	SDLVINQIKNRYKIINDNLNPSINEGDTFTMTKIFSDKLASISNE-----DNM	444	
DB	631	IDFILSGOKQYLEL-QIRADNISYNNLFYIT-----EDLKSANVAIGGIPAKKYLEHG	684	
QY	445	MFMIKITNLYKVGFPADVRSTNLSP-----	471	
DB	685	LP-----SEYRQDGTIPYVSTLNSGPDIMFQMKYKSLGRIGEVAHKONKLSVDNVL	740	
QY	472	GYVTGAYQDLMF---KONSTNIHLLEP-----LRNFEPKTKISOLT---BOEFTS	518	
DB	741	GVYASNNKDKSFMNLPVSVGINDITPDDESSWVRNNDINKILFEKINCHVPEKLPTS	800	
QY	519	LWSFNQARAKSOFEEYKGYFEGALGEDDNLDFAQNTVLDKDYVSKKILSSNMKTRKEYI	578	
DB	801	LY-----YEIDSRFFQGW---DNKSIKHVTEINKDLI--KDINLLLTSSNIDV	844	
QY	579	HTVLOQGDKIBYEASCNLFKDPYSSILYQKNIEGSETAYVYVYVADAEIKEDIRIPY	638	
DB	845	KLLIKL--DRELYAISKKI-----DNPLALSIRTLQOLANYVYTSNTPEPENTINFIY	896	
QY	639	QJNKRN-----IKLFTGHGKSEFTNTTANLVDTS-----LSSELET	677	
DB	897	DFYRKQDQLLSAHL-----FSRNDADTKIIVYNSVMKNVFLSEVISC	942	
QY	678	ILNLARADISPKYIENLLGCMFYSISAETYPGKLLKIKDKRVSELMPSISQDSI--	735	
DB	943	VLRSKKVD---SVINEN-----KKNLSKEDA--GALRDYAKLKKKELFSLMLDDGYKK	990	
QY	736	TVSANQVEVRINPEGKEILDHSGKWINKEESIIKDISKEYISFNP-----K	783	
DB	991	IITNAY---IKERDKL-----SGIYNIENSIISGHESFDIIRSNQHEWGLSTVEQFK	1042	
QY	784	ENKIIVKSKYLHELSTLQSIIRNANSOIDEKKVMLTECEINVAS-----NIDRQIVE	838	

US-10-282-122A-76865
Sequence 76865, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangseu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent in version 3.1
SEQ ID NO 76865
LENGTH: 4688
TYPE: PRT
ORGANISM: Ureaplasma urealyticum
US-10-282-122A-76865
Query Match 5.6%; Score 290.5; DB 12; Length 4688;
Best Local Similarity 18.8%; Pred. No. 8.5e-09;
Matches 252; Conservative 220; Mismatches 433; Indels 437; Gaps 65;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:39:30 ; Search time 15.375 Seconds
(without alignments)
3424.942 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1020
Perfect score: 5189
Sequence: 1 MNLVKAQIQKVVYKFRQ.....ITDASKVELYSTALDETID 1020

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/prodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/prodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/prodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/prodata/2/iaa/6C COMB.pcp.*
6: /cgn2_6/prodata/2/iaa/6D COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4101	79.0	2366	1	US-08-480-604A-10
2	4101	79.0	2366	2	US-08-405-496A-10
3	4101	79.0	2366	3	US-08-915-136-10
4	4101	79.0	2366	4	US-08-957-310-10
5	4101	79.0	2366	4	US-10-011-366-10
6	4101	79.0	2366	4	US-09-084-517-10
7	2791	53.8	2710	1	US-08-480-604A-6
8	2791	53.8	2710	2	US-08-405-496A-6
9	2791	53.8	2710	3	US-08-915-136-6
10	2791	53.8	2710	4	US-08-957-310-6
11	2791	53.8	2710	4	US-10-011-366-6
12	2791	53.8	2710	4	US-09-084-517-6
13	422.5	8.1	3169	4	US-09-453-702B-257
14	267	5.1	2807	4	US-09-543-681A-4980
15	257.5	5.0	2184	4	US-09-417-485D-6
16	250.5	4.8	2777	4	US-09-543-681A-6124
17	250.5	4.8	10182	4	US-09-134-001C-3159
18	239.5	4.6	3433	4	US-09-091-501B-10
19	239	4.6	3248	1	US-08-353-700-1
20	239	4.6	3248	5	PCT-US93-16216-1
21	221	4.3	2285	4	US-09-308-375-2
22	218	4.2	2662	4	US-09-595-684B-31
23	217	4.2	3696	4	US-09-134-001C-5080
24	215	4.1	956	4	US-09-134-001C-4452
25	211.5	4.1	2954	4	US-09-150-867-1
26	210.5	4.1	1151	4	US-09-134-001C-3242
27	209.5	4.0	1010	4	US-09-134-001C-5178

Sequence 8, Appli
Sequence 5821, Ap
Sequence 11, Appl
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 20, Appl
Sequence 190, App
Sequence 190, App
Sequence 190, App
Sequence 8, Appli

28 208.5 4.0 1786 3 US-08-973-462-8
29 208.5 4.0 2504 4 US-09-328-352-5821
30 205 4.0 3878 4 US-09-314-259-11
31 203.5 3.9 859 1 US-08-053-614-2
32 203.5 3.9 859 2 US-08-316-397B-2
33 203.5 3.9 859 1 US-09-034-306-2
34 203.5 3.9 859 3 US-09-259-437-2
35 203.5 3.9 859 5 PCT-US93-09782-2
36 203.5 3.9 1181 1 US-08-053-614-4
37 203.5 3.9 1181 1 US-08-316-397B-4
38 203.5 3.9 1181 2 US-09-034-306-4
39 203.5 3.9 1181 3 US-09-259-437-4
40 203.5 3.9 1181 5 PCT-US93-09782-4
41 202.5 3.9 1169 4 US-09-255-829-20
42 202 3.9 840 3 US-08-974-549A-190
43 202 3.9 840 4 US-09-402-181B-190
44 202 3.9 840 4 US-09-721-456-190
45 202 3.9 872 3 US-08-851-843A-8

ALIGNMENTS

RESULT 1
US-08-480-604A-10
; Sequence 10, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027

```
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-604A-10

Query Match 79.0%; Score 4101; DB 1; Length 2366;
Best Local Similarity 77.6%; Pred. No. 8.4e-238;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;

QY 1 MNLVNAQLQKQVYKRIQDEYVAILNALBEEYHNMSSSVVEKYKLDKINNLTDNYL 60
Db 1 MSLVNRKQLKQVAVRPTQDEYVAILDALBEEYHNMSENTEVVEKYKLDKINSLTDIYI 60
QY 61 NTKYKSGRNKALKKFKKYLTVKVELEKNNSLTPVEKNLHFWIGQINDTAINYQWKD 120
Db 61 DTYKSGRNKALKKFKKYLTVKVELEKNNSLTPVEKNLHFWIGQINDTAINYQWKD 120
QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNNTLESFRENLDNDFPDYKPFYKRMELIY 180
Db 121 VNSDYNVNFYDSNAFLNTLTKTIVESAINDTLESFRENLDNDFPDYKPFYKRMELIY 180
QY 181 DKQHFIDYKSGIEENFEFTIDNIITKYLSEYKDLALNKYIEBSLNKITANNNDI 240
Db 181 DKQKNFINKYKAQREENPELIDDIIVKTYLSNEYSKSIDELNTYIESLNKITQNSGNDV 240
QY 241 RNLEKFADELVLRYNQELVERWNLAAASDLIRLSMLKEDGGVLDVDIILPGIQPDLFKS 300
Db 241 RNFEFKNGESFNLYEQELVERWNLAAASDLIRLSALKETGGYLDVDMPLPGIQPDLFES 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDERVQSFESALSSKSKSEIF 360
Db 301 IEPSSVTVDPEWMTKLEAIMKYKEYIPGYTSEHFDMLDEEVQSFESVLASKSDKSEIF 360
QY 361 LPLDDIKVSPLEVKIAPANNVINQALISLKDSCSDIVINQIKNRYKILNDNLNPSINE 420
Db 361 SSLGMEASPLEVKIAFNKGIINQGLISVKDSCSNLIVKQIENRYKILNNSLNPAISE 420
QY 421 GTDFTNTWKIPSDKLASISNEDNMFMKITYNLKVGFPADVRSTINLSGGVYTGAYQD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMMELGKLYRVGFPDPVKTTINLSGPEAVAAAYQD 480
QY 481 LLMPKDNSTNHLLEPELANFEFPKTKISQTEQBITSLWSFNQARAKSQPEEYKKGYPF 540
Db 481 LLMPFKGSMNHLIEADLRNFEISKTNI-SQTEQEMASLWSFDDARAKAQPEEYKRYNPF 540
QY 541 CALGEDDNLDPQNTVLDKQVYKKSILSSMKTRNKEYIHYIYVQLQGDKISYEASCNLFSK 600
Db 541 SGLGEDDNLDPQNTVLDKQVYKKSILSSARSSRGYIHYIYVQLQGDKISYEACNLPFK 600
QY 601 DPYSILYQKNIEGSETAYYYVADAEIKEDKYRIPYQISNKNIKLTFIGHGKSFBNT 660
Db 601 IPYDSVLQKNLEDESEIAYVYVNGPGEIQTQEDKYKIPSIISDRPKIKLTFIGHGKDFNT 660
QY 661 DTFANLVDVSJSGSETITLNLAKADISPKYIETINLLGCMFSPYSISABETYPGKLLKIK 720
Db 661 DIFAGFDVDSLSTEIEAADIADLAKEDISPKSTEINLLGCMFSPYSINVEETYPGKLLKVK 720
QY 721 DAVSELMPSISODSITVSANQYEVINEBKGREILDHSGKWNKEESIIKDISKEYISF 780
Db 721 DKRISLMPISODSITVSANQYEVINEBGRREILDHSGWINKESIIKDISKEYISF 780
QY 781 NPKENKIIKSKYLHELSTLLOEIRNANSDDILEKKVMLTECEINVASNIDRQIVEGR 840
Db 781 NPKENKIIKSKYKHELSTLLOEIRNANSDDILEKKVMLTECEINVISNIDQIVEER 840
QY 841 IBEAKNLTSDSINYIKNEFKLIESDSLVDLKHQNGLDSDSHFISFEDISKTENGFRIRF 900
Db 841 IBEAKNLTSDSINYIKNEFKLIESISDALCDLQKQNELEDSHFISFEDISDEGFSIRF 900
QY 901 INKGTGNSIFETETKEIFSEYATHISKEISNKTIFDNVNGKLVKKVNDAAEVNTLN 960
Db 901 INKGTGESIFVETETKIFSEYANTHIEISIKGTIFDVTWNGKLVKKVNDTTHEVNTLN 960
QY 961 GAFFIQSLIEYNTKESLSNLSVAMKVQVYAQLFSTGLNTITTDASKVVELVSTALDETID 1020
Db 961 AAFFIQSLIEYNSKESLSNLSVAMKVQVYAQLFSTGLNTITTDAAKVVVELVSTALDETID 1020

RESULT 2
US-08-405-496A-10
; Sequence 10, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405.496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-496A-10

Query Match 79.0%; Score 4101; DB 2; Length 2366;
Best Local Similarity 77.6%; Pred. No. 8.4e-238;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;

QY 1 MNLVNAQLQKQVYKRIQDEYVAILNALBEEYHNMSSSVVEKYKLDKINNLTDNYL 60
Db 1 MSLVNRKQLKQVAVRPTQDEYVAILDALBEEYHNMSENTEVVEKYKLDKINSLTDIYI 60
QY 61 NTKYKSGRNKALKKFKKYLTVKVELEKNNSLTPVEKNLHFWIGQINDTAINYQWKD 120
Db 61 DTYKSGRNKALKKFKKYLTVKVELEKNNSLTPVEKNLHFWIGQINDTAINYQWKD 120
QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNNTLESFRENLDNDFPDYKPFYKRMELIY 180
Db 121 VNSDYNVNFYDSNAFLNTLTKTIVESAINDTLESFRENLDNDFPDYKPFYKRMELIY 180
QY 181 DKQHFIDYKSGIEENFEFTIDNIITKYLSEYKDLALNKYIEBSLNKITANNNDI 240
Db 181 DKQKNFINKYKAQREENPELIDDIIVKTYLSNEYSKSIDELNTYIESLNKITQNSGNDV 240
QY 241 RNLEKFADELVLRYNQELVERWNLAAASDLIRLSMLKEDGGVLDVDIILPGIQPDLFKS 300
Db 241 RNFEFKNGESFNLYEQELVERWNLAAASDLIRLSALKETGGYLDVDMPLPGIQPDLFES 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDERVQSFESALSSKSKSEIF 360
Db 301 IEPSSVTVDPEWMTKLEAIMKYKEYIPGYTSEHFDMLDEEVQSFESVLASKSDKSEIF 360
QY 361 LPLDDIKVSPLEVKIAPANNVINQALISLKDSCSDIVINQIKNRYKILNDNLNPSINE 420
Db 361 SSLGMEASPLEVKIAFNKGIINQGLISVKDSCSNLIVKQIENRYKILNNSLNPAISE 420
QY 421 GTDFTNTWKIPSDKLASISNEDNMFMKITYNLKVGFPADVRSTINLSGGVYTGAYQD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMMELGKLYRVGFPDPVKTTINLSGPEAVAAAYQD 480
QY 481 LLMPKDNSTNHLLEPELANFEFPKTKISQTEQBITSLWSFNQARAKSQPEEYKKGYPF 540
Db 481 LLMPFKGSMNHLIEADLRNFEISKTNI-SQTEQEMASLWSFDDARAKAQPEEYKRYNPF 540
QY 541 CALGEDDNLDPQNTVLDKQVYKKSILSSMKTRNKEYIHYIYVQLQGDKISYEASCNLFSK 600
Db 541 SGLGEDDNLDPQNTVLDKQVYKKSILSSARSSRGYIHYIYVQLQGDKISYEACNLPFK 600
QY 601 DPYSILYQKNIEGSETAYYYVADAEIKEDKYRIPYQISNKNIKLTFIGHGKSFBNT 660
Db 601 IPYDSVLQKNLEDESEIAYVYVNGPGEIQTQEDKYKIPSIISDRPKIKLTFIGHGKDFNT 660
QY 661 DTFANLVDVSJSGSETITLNLAKADISPKYIETINLLGCMFSPYSISABETYPGKLLKIK 720
Db 661 DIFAGFDVDSLSTEIEAADIADLAKEDISPKSTEINLLGCMFSPYSINVEETYPGKLLKVK 720
QY 721 DAVSELMPSISODSITVSANQYEVINEBKGREILDHSGKWNKEESIIKDISKEYISF 780
Db 721 DKRISLMPISODSITVSANQYEVINEBGRREILDHSGWINKESIIKDISKEYISF 780
QY 781 NPKENKIIKSKYLHELSTLLOEIRNANSDDILEKKVMLTECEINVASNIDRQIVEGR 840
Db 781 NPKENKIIKSKYKHELSTLLOEIRNANSDDILEKKVMLTECEINVISNIDQIVEER 840
QY 841 IBEAKNLTSDSINYIKNEFKLIESDSLVDLKHQNGLDSDSHFISFEDISKTENGFRIRF 900
Db 841 IBEAKNLTSDSINYIKNEFKLIESISDALCDLQKQNELEDSHFISFEDISDEGFSIRF 900
```

Db 61 DTYKSGRNKALKKFVLEVLKNNLTPVEKNLHFVWIGGQINDTAINYNQWKD 120
Qy 121 VNSDYTVKVDYVDSNAFLINTLKTIVESAATNTLESFRENLDPEFDYKFKRMEIY 130
Db 121 VNSDYVNVVDSNAFLINTLKTIVESAINDTLESFRENLDPRFDYKFKRMEIY 130
Qy 181 DKQGHFDYKVDYVDSNAFLINTLKTIVESAINDTLESFRENLDPRFDYKFKRMEIY 240
Db 181 DKQGHFDYKVDYVDSNAFLINTLKTIVESAINDTLESFRENLDPRFDYKFKRMEIY 240
Qy 241 RNLKFADEDLVRLYNOLVERNLAAASDILRAISMLKEDGVYLDVLPGLQDPLFKS 300
Db 241 RNLKFADEDLVRLYNOLVERNLAAASDILRAISMLKEDGVYLDVLPGLQDPLFKS 300
Qy 301 INKPDSTINTSWEMIKLEAIKMYKEYIPGYTSKNFMDLDEVOQSPESALSSKSDSEIF 360
Db 301 INKPDSTINTSWEMIKLEAIKMYKEYIPGYTSKNFMDLDEVOQSPESALSSKSDSEIF 360
Qy 361 LPDDIKVPLEVIAFANNVINOALISKDSYCSOLVINQIKNRYKILNDNLNPSINE 420
Db 361 LPDDIKVPLEVIAFANNVINOALISKDSYCSOLVINQIKNRYKILNDNLNPSINE 420
Qy 421 GTDFNTMTKIFSDKLASINEDNMFMKILNTNLYKGFADPVDSTINLSPGYVTGAYOD 480
Db 421 GTDFNTMTKIFSDKLASINEDNMFMKILNTNLYKGFADPVDSTINLSPGYVTGAYOD 480
Qy 481 LLMEKDNSTNHLPELRLNFEFKTKISQTEITSLMSFNQARAKSOFEEYKGYFE 540
Db 481 LLMEKDNSTNHLPELRLNFEFKTKISQTEITSLMSFNQARAKSOFEEYKGYFE 540
Qy 541 GALGEDNLPAQNTVDKDYVSKILSSMKTRNKEYIHYIVQLQGDKISYEASCNLFK 600
Db 541 GALGEDNLPAQNTVDKDYVSKILSSMKTRNKEYIHYIVQLQGDKISYEASCNLFK 600
Qy 601 DPYSILYOKNIESEFAYVYVADAEIKIDKIRIPYQISNKNKILFIHGKSEFNT 660
Db 601 DPYSILYOKNIESEFAYVYVADAEIKIDKIRIPYQISNKNKILFIHGKSEFNT 660
Qy 661 DTFANLVDLSSEIETILNAKADISPKYIEINLGCNMFYSISAEETYPCKLLKIK 720
Db 661 DTFANLVDLSSEIETILNAKADISPKYIEINLGCNMFYSISAEETYPCKLLKIK 720
Qy 721 DRVSELMPSISQDSITVSANQYEVRIEENEGKREILDHSGKWKINKEESIIKDISKEYISF 780
Db 721 DRVSELMPSISQDSITVSANQYEVRIEENEGKREILDHSGKWKINKEESIIKDISKEYISF 780
Qy 781 NPENKIIVSKYLHELSTLLOEIRNANSDDIDLEKKVMTLCEINVAISNIDROIVEGR 840
Db 781 NPENKIIVSKYLHELSTLLOEIRNANSDDIDLEKKVMTLCEINVAISNIDROIVEGR 840
Qy 841 IEEAKNLTSDSINIKNEFKLIESISDLYDLKHONGLDSDHSFISPEDISKTENGPRIF 900
Db 841 IEEAKNLTSDSINIKNEFKLIESISDLYDLKHONGLDSDHSFISPEDISKTENGPRIF 900
Qy 901 INKGTGSIPTETKEKIFSEVATHI SKEISNIDKTIPTDNVNGKLVKKNLDAAEVNTLN 960
Db 901 INKGTGSIPTETKEKIFSEVATHI SKEISNIDKTIPTDNVNGKLVKKNLDAAEVNTLN 960
Qy 961 SAFTIQSLIEYNTTKESLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 SAFTIQSLIEYNTTKESLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 3
US-08-915-136-10
; Sequence 10, Application US/08915136
; Patent No. 6290360
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.

APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-10

Query Match 79.0%; Score 4101; DB 3; Length 2366;
Best Local Similarity 77.6%; Pred. No. 8.4e-238;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;
Qy 1 MSLVNRKQLEKXVYKFRQDEYVAILDALEEYHNMSNTVVEKYLKDKDINLTNYL 60
Db 1 MSLVNRKQLEKXVYKFRQDEYVAILDALEEYHNMSNTVVEKYLKDKDINLTNYL 60
Qy 61 NYTKSGRNKALKKFVLEVLKNNLTPVEKNLHFVWIGGQINDTAINYNQWKD 120
Db 61 NYTKSGRNKALKKFVLEVLKNNLTPVEKNLHFVWIGGQINDTAINYNQWKD 120
Qy 121 VNSDYTVKVDYVDSNAFLINTLKTIVESAATNTLESFRENLDPEFDYKFKRMEIY 180
Db 121 VNSDYTVKVDYVDSNAFLINTLKTIVESAATNTLESFRENLDPEFDYKFKRMEIY 180
Qy 181 DKQGHFDYKVDYVDSNAFLINTLKTIVESAINDTLESFRENLDPRFDYKFKRMEIY 240
Db 181 DKQGHFDYKVDYVDSNAFLINTLKTIVESAINDTLESFRENLDPRFDYKFKRMEIY 240

```

QY 241 RNLEKFAEDLVRLYNQELVERWNLAAADILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFKNGESPNLYEQELVERWNLAAADILRISALKEIGWYLDVDMPLGQDPLFES 300
QY 301 INKPSITNTSWEMTKLEAIMKYKEYIPGYTSKMPDMLDEBQVRSFSAISSKDKSEIF 360
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKMPDMLDEBQVRSFSAISSKDKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINCALISLSDVCSDLVINOIKRYKILNDNLNPSINE 420
Db 361 SSLGDMEASPLEVKIAFANNVINCALISLSDVCSDLVINOIKRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKISDGLASISNEDNMFMKITYLVKVPAPDVRSTINLSGPGVVTGAYQD 480
Db 421 DNDFTNTTTFIDSINAEANADNGRFMMELGKYLVRGFPDPVKTINLSGPEAAAYQD 480
QY 481 LLMFKDGNSTNHLLEPELNFEPFKTKISOLTEQEIITSLWSFNQARAKSQFEYKGYFE 540
Db 481 LLMFKDGNSTNHLLEPELNFEPFKTKISOLTEQEIITSLWSFNQARAKSQFEYKGYFE 540
QY 541 GALGEDDMLDFAQNTVLDKDYVSKILSSMKTRKEYIHYIVOLQDKISYEASCNLFSK 600
Db 541 GSLGEDDMLDFAQNTVLDKDYVSKILSSMKTRKEYIHYIVOLQDKISYEASCNLFSK 600
QY 601 DYSSILYQKNTGSETAYVYVADAEIKEDKRIPIQISNKNRNIKTFIGHCKSEFNT 660
Db 601 TDYDVLFOKNTGSETAYVYVADAEIKEDKRIPIQISNKNRNIKTFIGHCKSEFNT 660
QY 661 DTFANLVDLSSETETILNLAADISPKYIEINLLGCNMFYSISABEYTPGKLLKIK 720
Db 661 DTFANLVDLSSETETILNLAADISPKYIEINLLGCNMFYSISABEYTPGKLLKIK 720
QY 721 DRVSELMPSISQDSITVSAQVYEVINBEGKREILDHSGKWINKEESIIDKISKEYISF 780
Db 721 DRVSELMPSISQDSITVSAQVYEVINBEGKREILDHSGKWINKEESIIDKISKEYISF 780
QY 781 NPKENKIVKSYLHSTLQIEIRNNSDDILEKKVMLTECEINVASNIDRQIVEGR 840
Db 781 NPKENKIVKSYLHSTLQIEIRNNSDDILEKKVMLTECEINVASNIDRQIVEGR 840
QY 841 IEAKNLTSDSINYIKNEFKLESISDLYDKHONGDDSHFISFEDISKTENGFRIRF 900
Db 841 IEAKNLTSDSINYIKNEFKLESISDLYDKHONGDDSHFISFEDISKTENGFRIRF 900
QY 901 INKETSNSIFETEKEIFESEYATHISKEISNIKOTIFDNVAGKLVKYNLDAHEVNTLN 960
Db 901 INKETSNSIFETEKEIFESEYATHISKEISNIKOTIFDNVAGKLVKYNLDAHEVNTLN 960
QY 961 SAFFQSLIEYNTVESLNSLVAMKVQVYAOPLSTGLNTITDASKVVELVSTALDETID 1020
Db 961 SAFFQSLIEYNTVESLNSLVAMKVQVYAOPLSTGLNTITDASKVVELVSTALDETID 1020

```

```

RESULT 4
US-08-957-310-10
; Sequence 10, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-957-310-10

```

```

Query Match 79.0%; Score 4101; DB 4; Length 2366;
Best Local Similarity 77.6%; Pred. No. 8.4e-238;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;

```

```

QY 1 MSLVNRKQLEKMANVRFRTQDEYVAIDLALSEYHNMSSENTVVEKYLKDKINSITDIYI 60
Db 1 MSLVNRKQLEKMANVRFRTQDEYVAIDLALSEYHNMSSENTVVEKYLKDKINSITDIYI 60
QY 61 NTYKSGRNRKALKKFKKYEITMEVLEKNNSLTPVEKNLHFIVIGQINDTAINYNQWMD 120
Db 61 DTYKSGRNRKALKKFKKYEITMEVLEKNNSLTPVEKNLHFIVIGQINDTAINYNQWMD 120
QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNNTLESFRENLDNPFYDYNKFKRMEIY 180
Db 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNNTLESFRENLDNPFYDYNKFKRMEIY 180
QY 181 DKQKHFIDYKQIENPEFIIDNTIKTYLSNEYSKDLKALNKYIEESLNKITAANGNDI 240
Db 181 DKQKHFIDYKQIENPEFIIDNTIKTYLSNEYSKDLKALNKYIEESLNKITAANGNDI 240
QY 241 RNLEKFAEDLVRLYNQELVERWNLAAADILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFKNGESPNLYEQELVERWNLAAADILRISALKEIGWYLDVDMPLGQDPLFES 300
QY 301 INKPSITNTSWEMTKLEAIMKYKEYIPGYTSKMPDMLDEBQVRSFSAISSKDKSEIF 360
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKMPDMLDEBQVRSFSAISSKDKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINCALISLSDVCSDLVINOIKRYKILNDNLNPSINE 420
Db 361 SSLGDMEASPLEVKIAFANNVINCALISLSDVCSDLVINOIKRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKISDGLASISNEDNMFMKITYLVKVPAPDVRSTINLSGPGVVTGAYQD 480
Db 421 DNDFTNTTTFIDSINAEANADNGRFMMELGKYLVRGFPDPVKTINLSGPEAAAYQD 480
QY 481 LLMFKDGNSTNHLLEPELNFEPFKTKISOLTEQEIITSLWSFNQARAKSQFEYKGYFE 540
Db 481 LLMFKDGNSTNHLLEPELNFEPFKTKISOLTEQEIITSLWSFNQARAKSQFEYKGYFE 540

```


QY 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHONGLDSDSHFISPDISKTEGPRIRF 900
DB 841 IEEAKNLTSDSINYIKNEFKLIESISDALCDLKQONELEDSDSHFISPDISKTEGPRIRF 900
QY 901 INKETSIFETETKEKEIFSEYATHISKEISNIDTIFDNVNGKLVKKVNDAAHEVNTLN 960
DB 901 INKETSIFETETKEKEIFSEYATHISKEISNIDTIFDNVNGKLVKKVNDAAHEVNTLN 960
QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAFPIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 6
US-09-084-517-10
Sequence 10, Application US/09084517
Patent No. 6613329
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: HAYSTACK, MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/084.517
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329.154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161.907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985.321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429.791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: OPND-01610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-084-517-10

Query Match 79.0%; Score 4101; DB 4; Length 2366;
Best Local Similarity 77.6%; Pred. No. 8, 4e-238;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;

QY 1 MNLVNAQOLQKVVYKFRIOEDYVAILNALVEEYHNMSESSVVEKYLLKLDINNLTDNYL 60
DB 1 MSLVNRKQLEKMANVRFETQEDYVAILDALEEYHNMSENTEVVEKYLLKLDINSLTDIYI 60
QY 61 NTKKSGKNGKALKKFKPEYLTVEVLELKNNSLTPVEKNLHPFWIGQINDTAINYNQWD 120
DB 61 DTKKSGKNGKALKKFKPEYLTVEVLELKNNSLTPVEKNLHPFWIGQINDTAINYNQWD 120
QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTTLESFRENLDNPEFQYKPKRMEIY 180
DB 121 VNSDYNVNFYDSNAFLINTLTKTIVESAINDTLESFRENLDNPRFDYKPKRMEIY 180
QY 181 DKQKHPIDYKSOIEBENPEFIDNIITKYLSNEYSKOLEALNKYIEBSINKITANGNDI 240
DB 181 DKQKFNIFYKAQREENPELIIDDIIVKTYLSNEYSKEIDELNTYTEESINKITQSGNDV 240
QY 241 RNLEKFADEDLVRLYNOELVERWNLAASDILIRISMLKEDGGVYLDVILFGIQDLPKS 300
DB 241 RNFEFKNGESFNLYEQELVERWNLAASDILIRISALKEIGMYLDVMDLFGIQDLPES 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTKNPFMDLDEEVQSFESALSCKSDKSEIF 360
DB 301 IEKPSVTVDFWMTKLEAIMKYKEYIPYTSHEFMDLDEEVQSFESVLSKSDKSEIF 360
QY 361 LFLDDIKVSPLEVKIAFANNVINOALISLKDSCDLVINQIKNRYKILNDNLNPSINE 420
DB 361 SSLGDMEASPLEVKIAFNKSGIINQGLISVSKSYCSNLIIVKQIENRYKILNLSNPAISE 420
QY 421 GTDFNTWKIPSDKLASISNEDNMFMKITYNLKVGFPADVRSTINISGPGVYTCAYQD 480
DB 421 DNDFTNTTTFIDSIIMAEANADNGFMELGKYLKLVGFPDPVKTINISGPEYAAAYQD 480
QY 481 LLMFKDNSNIHLPELNRNPEFPKTSQTEQRTSLWSFNQARAKSQPEEYKKGVE 540
DB 481 LLMFKEGSNHILHLEADJRNPEISKTNSQTEQMASLWSFDDARAKAQPEEYKRNVE 540
QY 541 GALGEDDNLDFQNTVLDKDYVSKKILSSMKTRKEYIHYIYVQLQDKISYEASCNLFPSK 600
DB 541 GSLGEDDNLDFQNTVLDKDYVSKKILSSARSSRGYIHYIYVQLQDKISYEACNLFPSK 600
QY 601 DPYSILYQKNTGSETAYYYYVADAETKEIDKYPYQISNKNRIKLTFTIGHGSEPT 660
DB 601 TPYDSVLFQKNTGSEIAYYYNPGDGEIQEIDKYPISIDRPKIKLTFTIGHGKDEFT 660
QY 661 DTFANLDVDSLSSEIETILNLAADISPKYIBINLLGCMFSPYSISABETPGKLLKIK 720
DB 661 DIFAGFDVDSLSTEIAAIDLAKEDISPKSIBINLLGCMFSPYSINVEETPGKLLKIK 720
QY 721 DRVSELMPISISODSITVSANOYEVRINEEGKEIILHSGKWKINKEESIIOISKEYISF 780
DB 721 DKISELMPISISQDSIIVSANQYEVNRINSEGRRELLDHSGEINKEESIIOISKEYISF 780
QY 781 NPKENKIIYKSKYLHELSTLLQEIERNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840
DB 781 NPKENKIIYKSKNLPESLTLQEIERNANSSDIELEKKVMLTECEINVSINIDTOIVEER 840
QY 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHONGLDSDSHFISPDISKTEGPRIRF 900
DB 841 IBEAKNLTSDSINYIKNEFKLIESISDALCDLKQONELEDSDSHFISPDISKTEGPRIRF 900
QY 901 INKETSIFETETKEKEIFSEYATHISKEISNIDTIFDNVNGKLVKKVNDAAHEVNTLN 960
DB 901 INKETSIFETETKEKEIFSEYATHISKEISNIDTIFDNVNGKLVKKVNDAAHEVNTLN 960
QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAFPIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 7
US-08-480-604A-6
Sequence 6, Application US/08480604A

Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6

Query Match 53.8%; Score 2791; DB 1; Length 2710;
Best Local Similarity 52.1%; Pred. No. 4.6e-159;
Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;

Qy 1 MNLVKAQLOQVYVKEPRIOEDVAVTLNLAEEVHNSSVVEKYLKLDINNTDNYL 60
Db 1 MSLISKEELKLAJ-SIRPRENEKYLTLNDELKNTLNNNENKYLQLKUNESIDVFM 59
Qy 61 NTKYKSGRNKALKKFKELTAMEVLELKNLSUTPVKKNLHPFWIGGQINDTAINYNQWKD 120
Db 60 NKYKTSRRNALSNLKDLKELIKNSNTSPVEKNLHFVWIGGVSDIALEYIQWAD 119
Qy 121 VNSDYTKVYVDSNAFLINTLKKTKTIVESATNTLESFRENLPDFTDYNKPKRWEIY 180
Db 121 VNSDYTKVYVDSNAFLINTLKKTKTIVESATNTLESFRENLPDFTDYNKPKRWEIY 180

Db 120 INAEYNIKLYDSEAFVNTLTKAIVSSSTTEALQLLEEEIQNFQDNMKFYKRMSEFIY 179
Qy 181 DKQHFIDYKYSQIEENPEFIIDNIITKLYSNEYSKOLEALNKYIEESLNKITANNNDI 240
Db 180 DRQKRFINYKYSQINKPTVPTIDDIKSHLVSEYNRDETVELSTYRSLKINSHGIDI 239
Qy 241 RNLEKPADEDLVRLYNQELVERWNLAASDLRLSMLEKEDGGVYLDVILFGIOPDLPKS 300
Db 240 RANSLFTEQELLNTYQSOLLNRGNLAASDIVRLALKNFGGVYLDVLMFGIHSIDLPKT 299
Qy 301 INKPSITNTSWEMIKLEAIKMYKEYEIPGYSKFNDFMLDEEVQSFESALSCKSKSEIF 360
Db 300 ISRPSSIGLDRWEMIKLEAIKMYKYYNNYTSNFDFKLDQQLKDNFKLIIESKSEKSEIF 359
Qy 361 LPLDDIKVSPLEVKIAPANNVINOALISLSDVINOIKRYKRIKILANDNLNPSINE 420
Db 360 SKLENINVSDELKIAFALGVSINOALISKOGSVLTNLVIEQVKNRYOFLNQHLPALLES 419
Qy 421 GTDFNTTMKIFSDKLASISNEDNMFMKITYNLKVGFPADVRSTINLSGGVTVTGAYQD 480
Db 420 DNNFTDTTKIFHDSLFNSATRENSMFLTKIAPYLQGVEMPEARSTISLSGGGAVASAYD 479
Qy 481 LLMFKONSTNIHLEPELNFEPFKTKISQTEGITSLSMFSNOARAKSQFEYKKGVE 540
Db 480 PINLQENTIEKTLKASDLIEFKFPENNLSQTEGIEINLSWFSFDOASAKYQEKYVRDYG 539
Qy 541 GALGEDNDLDFQONTVLDKDY-VSKKILSS--MKTRNKEYIHYIVOLGDDISSEASCNL 597
Db 540 GSLSEDNGVFNKNTALDKNYLLNKKIPSNVVEAGSNVYHYIIOQGGDISYEATCNL 599
Qy 598 FSKDPYSSILYQKNIEGSETAYVYVAD--AEIKEDKYRIPYOISNKRNIKLFIHGK 655
Db 600 FSKNPNKNSIIIQRM--NESAKSYFLSDGSEIIELEKNYRIPERLKNKEKVKVFIHGK 657
Qy 656 SEFNMTDPANLDDVLSSEIETILNLAKADISPKYIEINLAGCMFYSISAEETYPGL 715
Db 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLDISPKNVEVNLGCMFYSYDFNVEETYPGL 717
Qy 716 LLKIKDRVSELMPSISQDSITVSANQVEVRINEREGREILDHSGKWINKESSIIKD1SK 775
Db 718 LLSIMDKITSTLPDVNKNISITIGANQVEVRINEREGRELLAHSKWINKEEAIMSDLSK 777
Qy 776 EYISENPKENLIYKSKYLHELSTLLQEIIRNANSSDIDLEKVMLECEINVASNDRO 835
Db 778 EYIFPDSIDNKLKAKSNIPGLASISEDIKTLILDASVSPDTKFLNKLNISSIGDY 837
Qy 836 IVEGRIBEAKNLSDSINYIKNEFKLIESISDSLYDLKHONGLDDSHFISPEDISKTEG 895
Db 838 IYVEKLEPVKNIHNSIDDLIDEPNLENVSDLEYELKLANLDEKYLISFEDISKONST 897
Qy 896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNINKDTIFDNVNGKLVKVALDAHE 955
Db 898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGNLLDNIOQDHTSQ 957
Qy 956 VNTLNSAFFIOSLIEYNTTKESLNSLVAMKVQYQALFSTGLNTITDASKVVELVSTAL 1015
Db 958 VNTLNAAFFIOSLIDYSSNKDVLNLTSTSVKQYQALFSTGLNTIYDSIQLVNLISNAV 1017
Qy 1016 DETID 1020
Db 1018 NDTIN 1022

RESULT 8
US-08-405-496A-6
Sequence 6, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEUROTOXIN
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-6

Query Match 53.8%; Score 2791; DB 2; Length 2710;
Best Local Similarity 52.1%; Pred. No. 4.6e-159;
Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;

QY 1 MNLVAKAOLQKVVYKFIQDEYVAILNALALEEYHNMSESSVWEKYLKLDKNNLTNYL 60
DB 1 MSLISKELIKLAY-SIRPRENEYKTLTNDYENKLTNNENKYLQKLNESIDVFM 59
QY 61 NTYKSGRNKALKKPKFYLTWEVLELKNLSLTPVEKNLHFIWIGQINDTAINYQWKD 120
DB 60 NKYKTSRRNALSNLKDDILKEVILKNSNTSPVEKNLHFVWIGVSDIALEVIKQWAD 119
QY 121 VNSDVTVKVYDSNAFLNTLTKTIVSATNTLSPRENLDPEFDYKPKYKGMELIY 180
DB 120 INAEYIKLWYDSEAFVTLTKKAIVSESTEAQLLEEIQNFQDNMKFYKRMFEFIY 179
QY 181 DKQHFIDYKSOIENPEFIDNIKTVLSNEYSKDLEALNKYIEBSLNKITANNNDI 240
DB 180 DKQKFPYKQINKPVPTIDDIKSHLVSEYNDETVLESTVNTSLRKNINSHGDI 239
QY 241 RNLEKFAEDLVRLNQLVERWNLAASDILRISMLXEDGGVLDVDILPGIQDLPFKS 300
DB 240 RANSLFTEQELNLYSQELLNAGNLAASDILVLLALNFGGVYLDVMDLPGIHSDLFT 299
QY 301 INKPSITNTSWEMIKLBAIMKYKPIGYTSKNFMDLDEEVQSFESALSCKSKGRIF 360
DB 300 ISRPSIGLDREWEMIKLBAIMKYKPIYNNYTSNFDFKLDQQLKNFKLIIESKSEGRIF 359
QY 361 LPDLIDIKUSPLEVKAFAANNVINGALISLKDYSCLDVINQIKRYKILANDNLNPSINE 420
DB 360 SKLENLNSDLEIKIAFALGVSINGALISKQSYLTNLVIEQVKNRYQFLNQHLPATES 419

QY 421 GTDFNTTKMFIKSFDSKLASINEDNMFMIKITNYLVKGFAPDVRSTINISGPGVYTGAYOD 480
DB 420 DNNFTDTTKIFHDSLFNSGATENSFLTKIAPYLQVGFMPPEARSTISLSPGAYASAYD 479
QY 481 LLMFKDNSTNIHLLEPELRNFPFKTKISQITQEBITLSWNSQARAKSQPEYKKGYPE 540
DB 480 FINLQENTTEKTKASDLIEFKFPENNLSQLTEQBINSLWSPDQASAKYQPEKYVDYTG 539
QY 541 GAGEDNDLDFRQNTVLDKDY-VSKKILSS--MKTRNKEYHYIYQLOGDKLSYEASCNL 597
DB 540 GSLSEDNGVDFNKNTALDKNYLLNNKIPSNVVEAGSKNYHYIIQLQODDLSYATCNL 599
QY 598 FSKDPYSSILYQKNTIEGSETAYYYVAD--ABIKEIDKYRIPYQISNKNINIKLTFIGHGK 655
DB 600 FSKNPKNSIIQRNW--NESAKSYFLSDGSGSILELNKXVRIPELRKNKXKVTFIGHGK 657
QY 656 SEFNTDTFANLDVDSLSSEIETILNLAADISPKYIEINLGCNMFPSYSISABEYTPGKL 715
DB 658 DEFNTSEPARLSVDSLSNEISSFLDTIKLIDISPKNVEVNLGCNMFSDYFNVEETVPGKL 717
QY 716 LLKIKDRVSELMPISQDSITVSANQYEVRIINEEGKREILDHSGKWINKKEESIIKDISK 775
DB 718 LLSIMDKITSTLFDVNNKNSITIGANQYEVRIINESEKRELLAHSKWINKKEEAIMSLSK 777
QY 776 EYISFNPKENKIIKSKYLHELSTLLQELRNANSSDIDLEKKVMLTECEINVASNIDRQ 835
DB 778 EYIFFDSIDNKLKAKSKNIPGLASISEDIKTLILDASVSPDTKFINLNNKLNIESIGDY 837
QY 836 IVEGRIEEAKNLTSDSYINYNKNEFKLIESISDLYDLKHQNGLDSDHSFISFEDISKTENG 895
DB 838 IYVEKLEPVKNIHNSIDDLIDEFNLENVSDLYELKLNLDKYLISFEDISKNST 897
QY 896 FRIRFINKETGNSIFETKEIFSEYATHISEINIKDITFDVNGKLVKKNLDAHE 955
DB 898 YSVRFINKGSGSVVYETEKEIFSKYSEHITKEITKNSITTDVNGNLLDNIQDHTSQ 957
QY 956 VNTLASAFQISLJEYNTTKESISLVAMKYQVLAQFLSTGLNTITDASKVVELVSTAL 1015
DB 958 VNTLNAAFFIQSLIDYSSKNVDLNDLSTSVKQLAQLFSTGLNTIYDSIQLVNLISNAV 1017
QY 1016 DETID 1020
DB 1018 NDTIN 1022

RESULT 9
US-08-915-136-6
Sequence 6, Application US/08915136
Patent No. 6290960
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTIPOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136

[illegible]

[illegible]

Query Match	5.0%;	Score 257.5;	DB 4;	Length 2184;
Best Local Similarity	19.6%;	Pred. No. 4.7e-07;		
Matches 251;	Conservative 192;	Mismatches 446;	Indels 391;	Gaps 60;
Qy	10	QKQVVYKFIQDEYVAIINALEEY--HNMSGS--SVWEKYIKLXDI-----NNLTD	57	
Db	572	KKCIPIKLLGSKHNFKIFLQNVKVFLLFNKYKESPSLQNVMKNIQKVKNIPOKKISKYTKN	631	
Qy	58	NVL--NTYKKSGRNKAL-KKFEYLT---MEVLELKNNSLTPVEKNLHFHWIGSQIND	109	
Db	632	RILLKNIPONVYENKILHNNKKEIITNINDIKIYNNKQDNL-----NSNF-----KIKT	681	
Qy	110	TAIN-----YINQWQVDSYTVK-----VFYDSNAFLINTKKYIVESATNNLT	154	
Db	682	TLFNKLRRKYFNKKIKKINIAIQKHLMMRLIYFLNFYIMPLRRFFFTTKSEQTHUKTI	741	
Qy	155	-----ESFR-----ENLNDPEFDY--NKF-YRGRMEIIVDKOKPI	187	

Db	742	FDRKINWHTKISNFCPLVHQIFRPNKULKCRNEPQWDYQNMENYKKGKEKI--KTNKYI	799
Qy	188	DYKQSQIEENPEFIIDNIITKTYLSNEYSKOLEAL-----NKYTEBSLNKITANN--GNDI	240
Db	800	FIKQMKKSTNKCINNFKSSKCIPIKKKKKGLYNIIRHNNIFITKQMEKSKTNNLINKSI	859
Qy	241	RNLKFPAD--EDLVRLYNCEL---VERWNLAASDIILRISMLKEDGGVYLDVLLPGIQP	295
Db	860	DNLYKLEIKNKSVAPYIKKPYFKKFKYKFKYFALKQWYIHRMAKEE-----	904
Qy	296	DLPFSINKPDSITNTSWEMIKLEAMKY-----KEYIPGYTS-----KNFD	336
Db	905	---KS-----NIKLERAFKHPFIFAOEKEHLLKYSFHHFFQNRKINYKGRN	948
Qy	337	MLDEEVQ-----RSPESALUSSKDXSEIFLPDDIKVSPLEVKIAPANNS	381
Db	949	KLJHRIKNIILIKONGSIGVKNDKOTFLHLIKNSKNXN-----NNK	998
Qy	382	VINQALISIKDSYCSDLVINQIKRKYKIILNDLNPISINEGTDTMTKIFSDKLAISINE	441
Db	989	KCN-----KQYNN-----NNNNNNNNNNNNNNCKLUSNCKRYNRNN	1033
Qy	442	DNMMFMIKINYLVKGFADPVRSTINLSGPVYGTAYQDLL-----MFKDNSTNIHLBP	496
Db	1034	NNNK-----KAKONEANNIDD---SNLEKKKKIYIYKIIEKNFMWLNLSIN-HFISK	1086
Qy	497	ELRNFEPPTK-----ISQLEQBITSLWSNQAPAKSQPEYKKGIFEGALGEDDNL	549
Db	1087	KLRINWIPKKGRLRPLINLSTLNPVEIVKQRIPEILKSKSSE---FYFHNILNLERE	1142
Qy	550	DFAQNTVLDKDYVK-----KILLSMKTNR-----KE	576
Db	1143	KCKNIKRKRKYKKNQFNPNVSLNINCNFSLUKJGNMRHNNNSLFWNTLTKGTGIEKLKK	1202
Qy	577	YIHIV-----QLQGDKISYEASC-----NLFS--KDPYSS	605
Db	1203	WLHYKQWFWYKKMKKYIKKLNKNNKIYAYICIGDFSNCYEHINHNYLFKILKNFFDN	1262
Qy	606	I-----LYQKNIESEYAY-----YVADAABEIKEDKIRIPYOISN	642
Db	1263	INNFEIYLFKRSPLYNKLNNSFLSYYPWNVKSFGLHY-----IRNLRELIIKSHLD	1317
Qy	643	KRNITLFIHGKSEFN-----TDTFANLVD--SLSEIETILNLAkadispKYEINLL	696
Db	1318	NHFFLNQMFTKSKSLYIFADSYKSLQVDKRDIFMTIITVIRYYVLNIYFSIKEFKLN	1377
Qy	697	GCNMFYSISAEETYPQKLL-LKIKDRV-----SELMPSISODSITVSA	739
Db	1378	RKNIFYQIPOENQMGWYISVRDKKVENIKKWYLSMKKINHDEILESQNSSININN	1437
Qy	740	NOYEVRIINEGKBILDHSGKWINKESIIK-----DISKEYISFNPKENKIIVK	790
Db	1438	KNFMCINHEQDTE--EKGNTQNKEXHDIVIGIYNNSPDSTTTTHSSNNYKGNNIHVS	1494
Qy	791	SKY-----LHELSTLLOE-----IRNNANSSDIDLEKKVMLTECEINVASNIDRQIVE	838
Db	1495	GDKNDGILLHKGNNSMNECYVKDKCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	1553
Qy	839	GRIEAKNLTDSDI--NYIKNEFK--LIESIDSLYDLKHQNGLDD---SHFISFEDIS	890
Db	1554	--IKYHTKIDTDNKNHTYFANKFLNFDKLIISNIYGLPGFSLSNILCSLYIAYLD--	1609
Qy	891	KTENGFRIRFNKGTGNSIFETEKEIFSEYATHISKEISNIKOTIFDNVNGK-----	943
Db	1610	KNEEFOQLLYSEKQINNKYFLANGT---CNYFNLSLILRFDIDFLFITLKNKIKIFKN	1666
Qy	944	--LVKKVNLDAAEHVNTLNSAFFQSLI-----EYNTTKESLSNLVAMKY--	987
Db	1667	LILLKKI---WGSNINSSKTKIFKIPLIYKNDLLIYNFQNKYQKKYKIKNNKKIQSVRN	1723
Qy	988	-QVYAQLFSTGLNTITTDASK	1006
Db	1724	KRIHNLQVNAKKKGTYSVQK	1743

US 09 120 0100 0000 11-11-2004

Search completed: April 1, 2004, 16:45:56
Job time : 23.375 secs

Fri Apr 2 08:15:50 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 15:33:50 ; Search time 42.75 Seconds
(without alignments)
6741.484 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1020
Perfect score: 5189
Sequence: 1 MLNVNAQLQKMYVFRIQ.....ITDASKVELVSTALDEIID 1020

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4101	79.0	2366	2 AAR95011	Aar95011 C. diffi
2	4098	79.0	2366	2 AAW68388	Aaw68388 Clostridi
3	2792	53.8	546	2 AAY33700	Aay33700 C. sordei
4	2791	53.8	2710	2 AAR95016	Aar95016 C. diffi
5	2791	53.8	2710	2 AAW68387	Aaw68387 Clostridi
6	422.5	8.1	3169	6 ABO14444	Abol14444 E. coli 0
7	294	5.7	1254	2 AAR07503	Aar07503 Mercozite
8	294	5.7	1254	2 AAW24575	Aaw24575 Mercozite
9	290.5	5.6	4688	6 ABU48941	Abu48941 Protein e
10	287	5.5	1979	3 AAB18171	Aab18171 Plasmodiu
11	283	5.5	1516	3 AAB18195	Aab18195 Plasmodiu
12	279	5.4	1178	6 ABU24510	Abu24510 Protein e
13	276.5	5.3	1182	6 ABU25521	Abu25521 Protein e
14	275.5	5.3	1279	6 ABU24531	Abu24531 Protein e
15	272	5.2	980	3 AAB18294	Abu18294 Plasmodiu
16	270	5.2	1196	6 ABU24813	Abu24813 Protein e
17	267.5	5.2	1121	3 AAB18241	Aab18241 Plasmodiu
18	265	5.1	4620	6 ABU40997	Abu40997 Protein e
19	260	5.0	2485	3 AAB18172	Aab18172 Plasmodiu
20	258	5.0	1090	5 ABE54799	Abb54799 Lactococc
21	257.5	5.0	2184	4 AAE00425	Aae00425 P. faicip
22	256	4.9	903	6 ABU24404	Abu24404 Protein e
23	255	4.9	1875	6 ABR53560	Abr53560 Protein s
24	254.5	4.9	1053	6 ABM70125	Abm70125 Photorhab
25	253.5	4.9	1184	6 ABU25330	Abu25330 Protein e

26	252.5	4.9	1733	6 ABM68686	Abm68686 Photorhab
27	252	4.9	1948	5 ABP73774	Abp73774 Candida a
28	251.5	4.8	1712	3 AAB18205	Aab18205 Plasmodiu
29	250.5	4.8	10182	5 ABP38314	Abp38314 Staphyloc
30	249	4.8	1817	3 AAB18301	Aab18301 Plasmodiu
31	247.5	4.8	3547	6 ABM70634	Abm70634 Photorhab
32	246.5	4.8	1275	6 ABU25348	Abu25348 Protein e
33	245.5	4.7	1103	6 ABU48942	Abu48942 Protein e
34	245.5	4.7	1881	5 ABP73809	Abp73809 Candida a
35	245	4.7	1639	2 AAW54145	Aaw54145 P. faicip
36	245	4.7	1639	2 AAE29345	Aae29345 Plasmodiu
37	244.5	4.7	6641	5 ABU42656	Abu42656 Protein e
38	243.5	4.7	5024	4 AAG82935	Aag82935 S. epider
39	243	4.7	1928	6 ABR53023	Abr53023 Protein s
40	242.5	4.7	1345	3 AAB18236	Aab18236 Plasmodiu
41	242.5	4.7	1802	3 AAB18217	Aab18217 Plasmodiu
42	241.5	4.7	1089	5 ABB09491	Abb09491 AmEPV fir
43	241	4.6	1477	2 AAY19981	Aay19981 B. burgdo
44	241	4.6	1494	2 AAY19980	Aay19980 B. burgdo
45	241	4.6	1494	6 ABU19204	Abu19204 Protein e

ALIGNMENTS

RESULT 1
AAR95011
ID AAR95011 standard; protein; 2366 AA.
XX
AC AAR95011;

XX 16-OCT-2003 (revised)
DT 08-JUL-1996 (first entry)
XX
XX C. difficile toxin B.

XX Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin; diarrhoea;
KW therapy; diagnosis; vaccine.
XX
XX Clostridium difficile; VPI strain 10463 (ATCC 10463).

XX WO9612802-A1.
XX 02-MAY-1996.
XX
XX 23-OCT-1995; 95WO-US013737.
XX
XX 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.

XX (OPHI-) OPHIDIAN PHARM INC.
XX Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;
PI Firca JR;
XX
XX WPI; 1996-230603/23.
DR N-PSDB; AAT29247.
XX
XX Fusion proteins comprising non-toxin protein and part of toxin - useful
PT to form anti-toxins against Clostridium botulinum type A, and C.
PT difficile type toxins, and to treat C. difficile intoxication, partic.
PT diarrhoea.
XX
XX Claim 36; Page 313-323; 434pp; English.

XX Clostridium difficile VPI strain 10463 toxin B (AAR95011), the product of
CC the toxin B gene (AAT29247), is a cytotoxin associated with diarrhoeic
CC disease. It can be obtd. by expression in transformed E. coli hosts of
CC portions of DNA that together cover the entire toxin B gene. Toxin B, and
CC portions of it (see also AAR95012-13, AAR75371-72 and AAR95018), pref.
CC expressed as fusions to polyhistidine affinity tags or maltose binding

CC protein, are used to raise avian antibodies useful as antitoxins or
CC diagnostics, and in vaccine prodn. (Updated on 16-OCT-2003 to standardise
CC OS field)
XX
SQ Sequence 2366 AA;

Query Match	79.0%; Score 4101; DB 2; Length 2366;
Best Local Similarity	77.6%; Pred. No. 1.4e-213;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;	
QY	1 MNLVNAQLOKQVYKPIQDEYVAIILNALLEEYHNHNSSESVVEKYIKLKDINLTDNYL 60
DB	1 MSLVNRKQLEKMANVRFTQDEYVAIILNALLEEYHNHNSSESVVEKYIKLKDINSLTDIYI 60
QY	61 NTYKSGRNKALKKFKFKEVLTVEVLELKNNSLTPVEKNLHFIWIGQINDTAINYNQWKD 120
DB	61 DTYKSGRNKALKKFKFKEVLTVEVLELKNNSLTPVEKNLHFIWIGQINDTAINYNQWKD 120
QY	121 VNSDYTVKVFYDSNAFLNLTAKTIVESATNTLIESFRENLDNPFDPYKFKRMELIY 180
DB	121 VNSDYTVKVFYDSNAFLNLTAKTIVESATNTLIESFRENLDNPFDPYKFKRMELIY 180
QY	181 DKQKHFIDYKSOIEBENPEFIIDNIITKYLNEYSKDLEALNKYIEESLNKITANNNDI 240
DB	181 DKQKHFIDYKSOIEBENPEFIIDNIITKYLNEYSKDLEALNKYIEESLNKITANNNDI 240
QY	241 RNLEKFADEDLVRLNQELVERWNLAAASDILRISMLKEDGGVLDVLDLPGIQDLPFS 300
DB	241 RNLEKFADEDLVRLNQELVERWNLAAASDILRISMLKEDGGVLDVLDLPGIQDLPFS 300
QY	301 INKPSDITNTWEMIKLEAIMKYEYIPYTSKNFMDLDEEVQSFESALSKSKDKSEIF 360
DB	301 IEKPSDITNTWEMIKLEAIMKYEYIPYTSKNFMDLDEEVQSFESALSKSKDKSEIF 360
QY	361 LPDDIKVPLEVKTAFANNSVINQALISLSDSYCSDLVINQIKNRYKILANDNPSNE 420
DB	361 SSGDMEASPLEVKTAFANNSVINQALISLSDSYCSDLVINQIKNRYKILANDNPSNE 420
QY	421 GTDENMTMIKPSKCLASINENMMFMKIITNLYKVGAPDVRSTINLSGPOVYTGAYQD 480
DB	421 DNDENMTMIKPSKCLASINENMMFMKIITNLYKVGAPDVRSTINLSGPOVYTGAYQD 480
QY	481 LLMFKDNSTNIHLEPELNFPPPTKISQLETSLSWFSNQARAKSQFEYKKGYSF 540
DB	481 LLMFKGSGNHIHLEADLNFISKTNISQLETSLSWFSNQARAKSQFEYKKGYSF 540
QY	541 GALGEDNLDFAQNTVLDKDYVSKKILSNKTRNKEYIHVIYVQLQDKISYASCNLSK 600
DB	541 GSLGEDNLDFAQNTVLDKDYVSKKILSNKTRNKEYIHVIYVQLQDKISYASCNLSK 600
QY	601 DPYSSILYKQNIQEGSTAYVYVADAEIKEDKRYTPYQISKNKIKLTFIGHGKSEFT 660
DB	601 TPYDVLQKNIQEDSTAYVYVADAEIKEDKRYTPYQISKNKIKLTFIGHGKSEFT 660
QY	661 DTPANLDVSLSEIETILNLAADISPKYIEINLGCNMFYSISAEETYPKLLLIK 720
DB	661 DTPANLDVSLSEIETILNLAADISPKYIEINLGCNMFYSISAEETYPKLLLIK 720
QY	721 DRYSELMPISQDSITVSANQYEVRIINBEGREILDHSGKWNKESIIKDISKEYISF 780
DB	721 DKISELMPISQDSITVSANQYEVRIINBEGREILDHSGKWNKESIIKDISKEYISF 780
QY	781 NPENKIIVKSKYVHELSTLQIRNANSSDIDLEKKVMTCEINVASNIDRQIVEGR 840
DB	781 NPENKIIVKSKYVHELSTLQIRNANSSDIDLEKKVMTCEINVASNIDRQIVEGR 840
QY	841 IEEAKNLTSDSINVKNEPKLIESISDLYDLKHONGLDDSHFISPEDISCTENGPRIF 900
DB	841 IEEAKNLTSDSINVKNEPKLIESISDLYDLKHONGLDDSHFISPEDISCTENGPRIF 900
QY	901 INKETGNSIFETEKETFEYSYATHISKEISNIKDTIFDNYNGKLVKKNVLDAAHEVNTLN 960
DB	901 INKETGNSIFETEKETFEYSYATHISKEISNIKDTIFDNYNGKLVKKNVLDAAHEVNTLN 960

QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVOVYQAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAFPIQSLIEYNTTKESLSNLSVAMKVOVYQAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 2
AAW68388 standard; protein; 2366 AA.
XX
AC AAW68388;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium difficile toxin B.
XX
KW Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen;
KW pseudomembranous enterocolitis.
XX
OS Clostridium difficile.
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR N-PSDB; AAV30561.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 18; Page 241-249; 428pp; English.
XX
CC This is the amino acid sequence of Clostridium difficile toxin B, deduced
CC from the coding region (see AAV30561) of the toxin B gene. Fragments of
CC the toxin B gene have been cloned into various prokaryotic expression
CC systems, and assessed for the ability to express recombinant toxin B
CC protein in E. coli. It would be advantageous to use simple and
CC inexpensive prokaryotic expression systems to produce and purify high
CC levels of recombinant toxin B for immunisation purposes. The invention
CC specifically relates to recombinant proteins derived from Clostridium
CC botulinum toxins (see AAW68389-400) and their use as immunogens for the
CC production of vaccines and antitoxins
XX
SQ Sequence 2366 AA;

Query Match	79.0%; Score 4098; DB 2; Length 2366;
Best Local Similarity	77.5%; Pred. No. 2e-213;
Matches 791; Conservative 107; Mismatches 122; Indels 0; Gaps 0;	
QY	1 MNLVNAQLOKQVYKPIQDEYVAIILNALLEEYHNHNSSESVVEKYIKLKDINLTDNYL 60
DB	1 MSLVNRKQLEKMANVRFTQDEYVAIILNALLEEYHNHNSSESVVEKYIKLKDINSLTDIYI 60
QY	61 NTYKSGRNKALKKFKFKEVLTVEVLELKNNSLTPVEKNLHFIWIGQINDTAINYNQWKD 120
DB	61 DTYKSGRNKALKKFKFKEVLTVEVLELKNNSLTPVEKNLHFIWIGQINDTAINYNQWKD 120
QY	121 VNSDYTVKVFYDSNAFLNLTAKTIVESATNTLIESFRENLDNPFDPYKFKRMELIY 180
DB	121 VNSDYTVKVFYDSNAFLNLTAKTIVESATNTLIESFRENLDNPFDPYKFKRMELIY 180
QY	181 DKQKHFIDYKSOIEBENPEFIIDNIITKYLNEYSKDLEALNKYIEESLNKITANNNDI 240
DB	181 DKQKHFIDYKSOIEBENPEFIIDNIITKYLNEYSKDLEALNKYIEESLNKITANNNDI 240

Db 181 DKQKFNINYYKAQRENPELIIIDIVKTVLSNEYSKEIDELNTYIIEESLNKLTQNSGNDV 240
 Qy 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDPILGQIDPLFKS 300
 Db 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDPILGQIDPLFKS 300
 Qy 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQSFESALSKSDKSEIF 360
 Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKNFDMLEEVQSFESALSKSDKSEIF 360
 Qy 361 LPDDIKVSPLEVKIAPANNVINOALISLKDSCDVLVINOIKNEYKILNNDLNPSINE 420
 Db 361 SSLGDMEASPLEVKIAPANNVINOALISLKDSCDVLVINOIKNEYKILNNDLNPSINE 420
 Qy 421 GTDPTNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVDVSTINLSGPGVYTGAYOD 480
 Db 421 DNDPTNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVDVSTINLSGPGVYTGAYOD 480
 Qy 481 LLMFKDNSTNIHLEPELRNFEPKTKISQLETEQETTSLSWFSNQARAKQFEYKKGYPE 540
 Db 481 LLMFKDNSTNIHLEPELRNFEPKTKISQLETEQETTSLSWFSNQARAKQFEYKKGYPE 540
 Qy 541 GALGEDNLDPQAQTVLDKDVSKLISLMSKTRNKEYIHYIVQIGDKISYEASCNLFK 600
 Db 541 GSGEDDNLDPQAQTVLDKDVSKLISLMSKTRNKEYIHYIVQIGDKISYEASCNLFK 600
 Qy 601 DPYSIIYOKNIEGSETAYVYVADAEIKIDKVRIPQISNENKILTFIGHGKEFNT 660
 Db 601 TPYDSVLFQKNIEDESEIAYYVNPQDGEIQIDKIPISLSDRPKILTFIGHGKEFNT 660
 Qy 661 DTFANLDVSLSSIEIETILNAKADISPKVIEINLGCNMFYSISAEETYPGKLLIK 720
 Db 661 DIPAGFDVSLSTIEIAAIDLAKEDISPKVIEINLGCNMFYSISAEETYPGKLLIK 720
 Qy 721 DRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWINKESIIKDISKEYISF 780
 Db 721 DKISLMPISQDSIIIVSANQYEVRIINEEGKREILDHSGKWINKESIIKDISKEYISF 780
 Qy 781 NPKENKIIVKSKYHLSTLLOETRNANSSDIDLEKVMLTECEINVASNIDRQIVEGR 840
 Db 781 NPKENKIIVKSKYHLSTLLOETRNANSSDIDLEKVMLTECEINVASNIDRQIVEGR 840
 Qy 841 IEAKNTSDSINVIKNEFKLIESISLVDLKHONGLDSDHSPEDISKTENGFRIF 900
 Db 841 IEAKNTSDSINVIKNEFKLIESISLVDLKHONGLDSDHSPEDISKTENGFRIF 900
 Qy 901 INKETSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAHEVNTLN 960
 Db 901 INKETSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAHEVNTLN 960
 Qy 961 SAFFIQSLIEYNTTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
 Db 961 AAFPIQSLIEYNTTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
 RESULT 3
 ID AAY33700 standard, protein; 546 AA.
 XX AC AAY33700;
 XX DT 18-JAN-2000 (first entry)
 XX DE C. sordellii lethal toxin protein fragment.
 XX KW Lethal toxin; immunotoxin; antitumor; Glucosyl transferase; glycosylate;
 XX GW GTPase; Ras; epidermal growth factor stimulated MAP-kinase; inhibition;
 XX KW signalling pathway; cell-specific toxin; treatment; cancer.
 OS Clostridium sordellii.
 XX DE 19802569-A1.
 XX PN
 XX

PD 09-SEP-1999.
 XX 23-JAN-1998; 98DE-01002569.
 XX 23-JAN-1998; 98DE-01002569.
 XX (UYFR-) UNIV FREIBURG ALBERT-LUDWIGS.
 XX Aktories K, Hofmann F;
 WPI; 1999-509323/43.
 DR N-PSDB; AAZ23800.
 XX New fragment of the lethal toxin from Clostridium bacterium, useful for
 PT treating cancer.
 PT Claim 1; Page 7-9; 14pp; German.
 PS This invention describes a novel fragment (I) of the lethal toxin (LT) of
 XX Clostridium sordellii which has antitumor activity. (I) is a Glucosyl
 CC transferase that glycosylates, and thus inactivates, GTP(guanine
 CC triphosphate)ases, particularly Ras (an oncogenic product overexpressed
 CC in many tumors), resulting in inhibition of epidermal growth factor
 CC stimulated MAP-kinase signalling pathways. (I), particularly in the form
 CC of immunotoxins, are used as cell-specific toxins, particularly for
 CC treating cancer. When included in immunotoxins, (I) can be targeted to
 CC selected cells. Compared with the complete LT, (I) is smaller, so enters
 CC cells more easily, resulting in greater toxicity in the cytosol, is less
 CC likely to induce formation of (neutralizing) antibodies, and is more
 CC active than the holotoxin. This sequence represents the lethal toxin
 CC fragment described in the invention
 XX
 SQ Sequence 546 AA;
 Query Match 53.8%; Score 2792; DB 2; Length 546;
 Best Local Similarity 99.5%; Pred. No. 2.e-143;
 Matches 543; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 MNLVNAQLQKVVYKFRIOEDYVAILNLEEEYHNMSESSVVEKYLKLDKINLTDNYL 60
 Db 1 MNLVNAQLQKVVYKFRIOEDYVAILNLEEEYHNMSESSVVEKYLKLDKINLTDNYL 60
 Qy 61 NTYKSGRNKALKKPKFVLTWEVLKNNSLTPVEKNLHFHWIGQINDTAINYNQWKD 120
 Db 61 NTYKSGRNKALKKPKFVLTWEVLKNNSLTPVEKNLHFHWIGQINDTAINYNQWKD 120
 Qy 121 VNSDVTYKVFYDSNAFLNLTAKTIVESATNTLTLESFRENLDNDFYKFKRKEWIIY 180
 Db 121 VNSDVTYKVFYDSNAFLNLTAKTIVESATNTLTLESFRENLDNDFYKFKRKEWIIY 180
 Qy 181 DKQKHFIDYKSOIEENPEFIIDNIIKTVLSNEYSKDLKALNKYIEESLNKLTANNNDI 240
 Db 181 DKQKHFIDYKSOIEENPEFIIDNIIKTVLSNEYSKDLKALNKYIEESLNKLTANNNDI 240
 Qy 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDPILGQIDPLFKS 300
 Db 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDPILGQIDPLFKS 300
 Qy 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQSFESALSKSDKSEIF 360
 Db 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQSFESALSKSDKSEIF 360
 Qy 361 LPDDIKVSPLEVKIAPANNVINOALISLKDSCDVLVINOIKNEYKILNNDLNPSINE 420
 Db 361 LPDDIKVSPLEVKIAPANNVINOALISLKDSCDVLVINOIKNEYKILNNDLNPSINE 420
 Qy 421 GTDPTNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVDVSTINLSGPGVYTGAYOD 480
 Db 421 GTDPTNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVDVSTINLSGPGVYTGAYOD 480
 Qy 481 LLMFKDNSTNIHLEPELRNFEPKTKISQLETEQETTSLSWFSNQARAKQFEYKKGYPE 540
 Db 481 LLMFKDNSTNIHLEPELRNFEPKTKISQLETEQETTSLSWFSNQARAKQFEYKKGYPE 540

QY	541	GALGED 546	
Db	541	GALGED 546	
RESULT 4			
AAR95016			
ID	AAR95016	standard; protein; 2710 AA.	
XX	AC	AAR95016;	
XX	DT	16-OCT-2003 (revised)	
DT	08-JUL-1996	(first entry)	
XX	DE	C. difficile toxin A.	
XX	KW	Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin; vaccine;	
KW	diarrhoea; diagnosis; therapy.		
XX	OS	Clostridium difficile; VPI strain 10463 (ATCC 10463).	
XX	PN	W09612802-A1.	
XX	PD	02-MAY-1996.	
XX	PF	23-OCT-1995; 95WO-US013737.	
XX	PR	24-OCT-1994; 94US-00329154.	
PR	16-MAR-1995; 95US-00405496.		
PR	14-APR-1995; 95US-00422711.		
PR	07-JUN-1995; 95US-00480604.		
XX	PA	(OPHI-) OPHIDIAN PHARM INC.	
XX	PI	Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;	
PI	Firca JR;		
XX	WPI; 1996-230603/23.		
DR	N-PSDB; AAT29248.		
XX	XX	Fusion proteins comprising non-toxin protein and part of toxin - useful	
PT	to form anti-toxins against Clostridium botulinum type A, and C.		
PT	difficile type toxins, and to treat C. difficile intoxication, partic.		
PT	diarrhoea.		
XX	PS	Claim 63; Page 290-302; 434pp; English.	
XX	XX	Clostridium difficile VPI strain 10463 toxin A (AAR95016), the product of	
CC	the toxin A gene (AAT29248), is a potent cytotoxin that plays a direct		
CC	role in damaging gastrointestinal tissues and is associated with		
CC	diarrhoeic disease. It can be obtd. by expression in transformed E. coli		
CC	hosts of portions of DNA that together cover the entire toxin A gene.		
CC	Toxin A, and portions of it (see also AAR95014-15 and AAR95017), pref.		
CC	expressed as fusions to polyhistidine affinity tags or maltose binding		
CC	protein, are used to raise avian antibodies useful as antitoxins or		
CC	diagnostics, and in vaccine prodn. (Updated on 16-OCT-2003 to standardise		
CC	OS field)		
XX	XX	Sequence 2710 AA;	
SQ	Query Match	53.8%; Score 2791; DB 2; Length 2710;	
	Best Local Similarity	52.1%; Pred. No. 2e-142;	
	Matches	534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;	
QY	1	MNLVNAQLOQVYKFRIOEBYVAILNALAEYHNMSSSVVEKYLKDKDINNUTDNYL 60	
Db	1	MSLISKEELIKLAY-SIRPRENEYKILTNLDYNNKLTNNENKYLQKLKLNESIDVFM 59	
QY	61	NTYKSGRNKALKKFYEYLTWELKNSLPVEKNLHFMTGGGINDTAINYNQKD 120	
Db	60	NKYKTSNRNALGNLKDKILKEVILKNSGTSPEKNLHFVWIGVEVSDIALEYIKQMD 119	

QY	121	VNSDYTVKVFYDSNAFLINTLTKKTIVESATNNTLESFRENLDPEFDYKFKRMELIY 180
Db	120	INAEYNIKWYDSEAFVNTLTKAIVESSTTALQLEEBEIQNPFQDNMKFKRMELIY 179
QY	181	DKQKHFIDYKSOIEENPEFIIIDNIITKLYLSNEYSKOLEALNKYEESLNKITANNGNDI 240
Db	180	DRQKGFINYKSOINKPTVPTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSNHDGI 239
QY	241	RNLEKFADEDLVRLYNQBELVERWNLAASDILRLISMLKEDGCVYLDVILPGIOPDLFKS 300
Db	240	RANSLFTQELLNIIYSQELLNRGNLAASDIVRLIALKNFGGVYLDVDMVLFQIHSDLFKT 299
QY	301	INKPDSITNTSWEMIKLEAIMKYEYIPGYTSKNFMDLDEEVQSFESALSCKSKSEIF 360
Db	300	ISRPSSIGLDRWEMIKLEAIMKYKYINNYTSENFDKLDQQLKONFKLIIESKSEIF 359
QY	361	LPDDIKVSPLEVKJAFANNSVINQALISLKSYSQDLVINOIKRKYILNDLNPSINE 420
Db	360	SKLENLNSDLEIKIAFALGVSINQALISKQSYLTNLVIEQVKNRYQFLNQHLPAYES 419
QY	421	GTDENTTKIPSDKLASISNEDNNMFMIKITNLYKVGFPAPVRSITNLSGPGVYTCAYQD 480
Db	420	DNNFTDTTKIPHDSLFNSATAENSFLTKIAPYLVGFMPPEARSTISLGPAYASAYD 479
QY	481	LLMFKDNSTNIHLPELRNFPFKTKISQLTEQBITSLWSFNQARAKSQPEYKKGYPF 540
Db	480	FINLQENTIEKTLKASDLIEFFPENNLSQLTEQINSLSWSPDQASAKYQFEKYVRYDTG 539
QY	541	GALGEDNDLDFACNTVLDKDY-VSKKLSS--MKTRNKEYIHYIVOLQDKLISYASCNL 597
Db	540	GSLSEDNGVDFNKNTALDKNYLLNNKIPSNVVEAGSKNYVHYIIQLQDDDISYEATCNL 599
QY	598	FSKDPYSSILYQKNTGSETAYVYVAD--ABIKEIDKYRIPYQISNKENIKLTFIGHCK 655
Db	600	FSKNPKNSIIQRNM--NESAKSYFLSDDGESILELNKYRIPERLNKKEKVKVTFIGHCK 657
QY	656	SEBNTDTRANLDVDSLSSBIETILNLAKADISPKYIEINLLGCNMFSYSISAEETYPGKL 715
Db	658	DEFTSEFARLSVDSLSSNISSEFLDTIKLIDISPKVVEVNLGCNMFSYDFNVEETYPGKL 717
QY	716	LLKIKDRVSELMPFSISQDSITVSANQYVRINEEGREILDHSGKWINKESIIKDISK 775
Db	718	LLSIMDKITSTLPDVNKNISITIGANGYEVRIINSEGRKELLAHSGKWINKEEAIMSDLSSK 777
QY	776	EYTSFNPKNKIIVKSKYLHELSTLLOETIRNNANSDDIDLEKKVMLTECEINVASNIDRQ 835
Db	778	EYIFFSDIDNKLKAKSKNIPGLASISEDIKTLILDASVSPDTKFIILNNLKMAIESIGDY 837
QY	836	IVEGRTEEAKNLTSDSINVIKNEFKLIESISLSYDLKHQNGLDSDHSFISFEDISKTENG 895
Db	838	IYYEKLPEPVKNIITHNSIDDLIDFNLENVSDLEYELKLNLDLDEKYLISFEDISKNNST 897
QY	896	FRIRFINKETGNSIFITEKEIPSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAHE 955
Db	898	YSVRFINKNGESVYVETEKEIPSKYSEHITKEISTIKNSIITDVNGNLLDNQLDHTSQ 957
QY	956	VNTLNSAFFTQSLIEYNTTKESLSNLSVAMKQVVAQLFSTGLNTITDASKVVELVSTAL 1015
Db	958	VNTLNAAFFIQSLIDYSSNKKVLDNLSTSVKQVLAQLFSTGLNTIYDSIQLVNLISNAV 1017
QY	1016	DETID 1020
Db	1018	NDTIN 1022
RESULT 5		
AAW68387		
ID	AAW68387	standard; protein; 2710 AA.
XX	AC	AAW68387;
XX	DT	07-DEC-1998 (first entry)
XX		

DE Clostridium difficile toxin A.
XX
KW Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen;
KW pseudomembranous enterocolitis.
XX
XX Clostridium difficile.
XX
XX WO9808540-A1.
XX
XX 05-MAR-1998.
XX
XX 28-AUG-1997; 97WO-US015394.
XX
XX 28-AUG-1996; 96US-00704159.
XX
XX (OPHI-) OPHIDIAN PHARM INC.
XX
XX Williams JA, Thalley BS;
XX
XX WPI; 1998-230234/20.
XX
XX N-PSDB; AAV30560.
XX
XX Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
XX Example 15; Page 220-230; 428pp; English.
XX
XX This is the amino acid sequence of Clostridium difficile toxin A, deduced
CC from the coding region (see AAV30560) of the toxin A gene. Toxin A is a
CC potent cytotoxin that plays a direct role in damaging gastrointestinal
CC tissues. Severe cases of C. difficile intoxication result in
CC pseudomembranous colitis. This would be prevented by neutralising the
CC effects of toxin A in the gastrointestinal tract. Examples are provided
CC of the production of recombinant C. difficile toxin A in host cells and
CC of the in vivo neutralisation of toxin A by antibodies against
CC recombinant toxin A polypeptides. The invention specifically relates to
CC recombinant proteins derived from Clostridium botulinum toxins (see
CC AAV6389-400) and their use as immunogens for the production of vaccines
CC and antitoxins
XX
XX Sequence 2710 AA;
XX
XX Query Match 53.8%; Score 2791; DB 2; Length 2710;
XX Best Local Similarity 52.1%; Pred. No. 2e-142;
XX Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;
XX
XX 1 MNLVKAQLQGVVTKFRIODEYVAIINALEEYHNMSESVVEKYLKDKDINNLDNYL 60
XX 1 MSLISKELIKLAY-SIRPRENEVKTLTNDLDEYNKLTNNENKYLQKLKLNESIDVFM 59
XX
XX 61 NTKYKSGRNKALKKPKFYLTMVEVLKXNLSLTPVERKLNHFIIWGGQINDTAIYNQWKD 120
XX 60 NKYKTSRNRALSNLKXDKILKEVLIKNSNTSPVEKNLHFVWIGGEVSDIALEVIKQWAD 119
XX
XX 121 VNSDYTKVFDVSNAPLINTLTKTIVESAATNTLESFRENLDPEFPYKFKRMELIY 180
XX 120 INAEYNIKLWYDSEAFVNTLTKAIVESSTTEALQLLEBEEIQNPQPDNMFKYKRMFEFY 179
XX
XX 181 DKQHFIDYKKSQJEEENPEFTIDNIITKYLNSNEYSKDLALNKYIBESLNKTIANNNGDI 240
XX 180 DRQKRFINYKSKQNKPTVPTIDIIKSHLVSEYNRDETLESYRTNLSLRKINSNGIDI 239
XX
XX 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVLDLPGIQDILFKS 300
XX 240 RANSLFTEQELLNYSQELLNARGNLAAASDIVRLALKNFGVYLDVDMPLGHSIDLFKT 299
XX
XX 301 INKPDSTINTSWEMIKELAIMKYKEYIPGYTSKPNFMDLDEVRQSFESALSSKSDKSEIF 360
XX 300 ISRPSSIGLDREWEMIKELAIMKYKYKYNNTSENFDKLDQOLKDNFKLIIESKSEKSEIF 359
XX
XX 361 LPDDIKVSPLEVIKAFANNSVINQALISLSDKSCVDLVIQIKYKILNDNLNPSINE 420
XX
XX

DB 360 SKLENLVSDLEIKIAFALGVSIVNQALISKQGSYLTNLVIEQVKNRYQFLNQHLPALIS 419
QY 421 GTDNTMTKIPSKYKLASINEDNMFMKLTNYLVKGFADPVESTINLSGPGVYTGAYOD 480
DB 420 DNNFTDTTKIFHDSLNSATAENSMTLKIAPLYQVGFMEARSTISLSGPGAYASAYD 479
QY 481 LLMEFKDSTNHLLEPRLNRPFPKTKISQITSEITSLWSFNQARAKSOFEEYKGYFE 540
DB 480 FINQENTIEKTLKASDLIEFPENNLSQLTQEIINLSWFSFQASAKYQFEKVRDYG 539
QY 541 GALGEDNDLDFQNTVLDKDY-VSKYLSS--MKTRNKEYIHYIVQLQGDKISYEASCNL 597
DB 540 GSLSEDNGVDFNKVNTALDKYLLNNKIPSNVVEAGSKYVHYIQLQGGDISYEATCNL 599
QY 598 FSKDPYSILYQKNIEGSETAYVYVAD--AEIKEDIKYIPQVQISNKNRIKLTFFIGHGK 655
DB 600 FSKNPKNSIIQRNM--NESAKSYFLSDGSGESILELNKYRIPERLKNKEKVKVTFIGHGK 657
QY 656 SEFNTDTFANLDVDSLSESEITILNLAADISPKYIEINLLGCMFYSISAEETYPGKL 715
DB 658 DEFNTSEFARLSVDSLSENEISSFLDTIKLDISPKNVEVNLGCMFYSYDFNVEETYPGKL 717
QY 716 LLKIKDRVSELMPSISODSITVSANQVEVRINEREGKEIILHSGKWINKEESIIKDISK 775
DB 718 LLSIMDKITSTLPDVNKNSTIIGANQVEVRINSEKRELLAHSKWINKEEAINMSDLSK 777
QY 776 EYISFNPKNKIIIVKSKYLHELSTLQEIIRNANSSDIDLEKKVMLTECEINVASNIDRQ 835
DB 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLILDASVDPDTKEILNKLNISSIGDY 837
QY 836 IVEGRIEAKNLTSDSINYKNKPKLIESDSYDIKHQNGLDSDHIFISPEDISKTEG 895
DB 838 IYVEKLEPVKNIHNSIDDLIDFENLLENVSDLEYELKLNLDKYLISPEDISKNST 897
QY 896 FRIPINKETGNSFIETEKIEFSEYATHISKEISNKTIFDNVNGKLVKVNLDAAHE 955
DB 898 YSVRFINKSGESVYVETEKIEFSKSEHITKEISTKNSIITDVNGLLDNILQDHTSQ 957
QY 956 VNTLNSAFFIOSLIEYNTTKESLNSLVAMKVQVYQALFSTGLNTITDASKVVELVSTAL 1015
DB 958 VNTLNAFFIOSLIDYSSNKVDLNDLSTSVKQVYQALFSTGLNTIYDSIQVLNLSNAV 1017
QY 1016 DETID 1020
DB 1018 NDTIN 1022
XX
XX ABO14444 standard; protein; 3169 AA.
XX ABO14444;
XX
XX 23-OCT-2003 (revised)
DT 21-AUG-2003 (first entry)
XX
XX E. coli 0157 clostridial cytotoxin-like protein o3169.
XX
XX OZID; acute haemorrhagic colitis; haemolytic uraemic syndrome;
KW food poisoning; clostridial toxin-like protein.
XX
XX Escherichia coli; strain 0157:H7.
XX
XX US2003023075-A1.
XX
XX 30-JAN-2003.
XX
XX 01-APR-2002; 2002US-00114170.
XX
XX 04-DEC-1998; 98US-0110955P.
DB 03-DEC-1999; 99US-00453702.
XX
XX (BLAT/) BLATTNER F R.
PA

PA (BURL/) BURLAND V D.
 PA (PERN/) PERNA N T.
 PA (PLUN/) PLUNKETT G.
 PA (WELC/) WELCH R.
 XX
 PI Blattner FR, Burland VD, Perna NT, Plunkett G, Welch R;
 XX
 XX WPI; 2003-479497/45.
 DR N-PSDB; ACD19243.
 XX
 XX New DNA sequences from *Escherichia coli* strain O157:H7, useful for
 PT detecting *E. coli* O157:H7 in a sample, or in designing diagnostic probes
 PT which can be used to distinguish strain O157:H7 from strain K12 using
 PT molecular techniques.
 XX
 XX Disclosure; SEQ ID NO 257; 33pp; English.
 XX
 CC The invention relates to an isolated DNA molecule comprising an *E. coli*
 CC strain O157:H7 sequence selected from a clostridial cytotoxin-like gene,
 CC a urease gene cluster, a RIX toxin-like gene cluster, a locus of
 CC enterocyte effacement and 2 genes from its associated lymphocytic phage
 CC 933W (a putative serine/threonine kinase and a tail fibre gene), *E. coli*
 CC O157:H7 can cause food poisoning, specifically acute haemorrhagic colitis
 CC (which can develop into haemolytic uraemic syndrome). Also included are
 CC an isolated DNA molecule comprising a nucleotide sequence identical to at
 CC least 25 contiguous nucleotides contained in DNA sequences selected from
 CC ACD18988-ACD19242 (being 255 *E. coli* O157 DNA sequences which are not
 CC found in *E. coli* K12), a recombinant DNA construction comprising the DNA
 CC above and a method for detecting *E. coli* O157:H7 (ATCC 43895) in a sample
 CC (or distinguishing between O157 and K12) using a probe derived from one
 CC of the 255 sequences. The DNA sequences are useful in detecting *E. coli*
 CC O157:H7 in a sample, for the early diagnosis of humans and livestock
 CC infected with O157:H7, and in designing diagnostic probes which can be
 CC used to distinguish strain O157:H7 from strain K12 using molecular
 CC techniques. The present sequence is the clostridial toxin-like protein.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO at seqdata.uspto.gov/sequence.html?DocID=20030023075 (Updated on 23
 CC -Oct-2003 to standardise OS field)
 XX
 SQ Sequence 3169 AA;
 Query Match 8.1%; Score 422.5; DB 6; Length 3169;
 Best Local Similarity 21.0%; Pred. No. 7.5e-14;
 Matches 258; Conservative 197; Mismatches 450; Indels 323; Gaps 54;
 14 YVKFRIT---QEVEYVAINAL--EEYHNMSSESVVEYKLVKLDINNLDNLTNYLYKSGR 68
 160 YIKIRTRGAEQDTTITQSLIINELLNGVDRNTI-PPQKISELNDIHSYENMQIKNSR 218
 69 N--KALKPKFYLTMEVLELKNS-----LTPVEKNLHFTWIGQ 106
 219 KGTEILVKGELLSSLLNDKNGKQLSDNASKIINLLGIEYQSHKVDIEPTTHAVWAGA 278
 107 INDATNINYNQKDVNSDYTKVFDNSAF-----LINTLKT----- 144
 279 PDNTFSYITAFNLTYKYDYTLWIDPNAFGAAGKSGILKNAMYAMRURTNPHLAE 338
 145 -----IVESATNNTLE--SPRENLDPEFY-----NKFYRKMEIIVDKQKHF 187
 339 ENNEVLTKQNTQNEIEPEKTEERLKELENYKSLTSEKFNVPFLESIGMQDNYF 398
 188 DYYKSQIENPEFI-----IDNIK--TYLSNEYSKDLALNKYIEESLNKITTANGN-- 238
 399 TYCISNGISNTDDISRLDFTNLVLSPEVQDFKSTVEKNKRD-DLLKNNTISQKFGDRF 458
 239 ---DIRNLEKFADEDLVRNLQNELVERWNLAAASDILRISMLKEDGGVLDVLDILPGIQP 295
 459 QLRDINTLESFKKPQDFFYQOEMLRWNYAAASDQVRINILKEYGGIYTDILPAYSD 518
 296 DLFKSKINKPDSINTSWEMIKLEAIMKYK--EYIPG--YTSKNFMDLDEVOQRSFESALS 351
 519 KVSQIINE-KSDDKRRFEDLKURRISESILSIKGEKYSIKH-DGLDETTLNQLNLI- 575

352 SKSDKSEIFLDDDIKVSPLVKIAPANNVI-----NQALISIKDSYC 395
 576 SEIEK-----LTIDY-FKPEVETKVVDRDFKPKYQKWENTWIRGNMNFMLTKGSKC 630
 396 SLVINOIKVRYKILNDNLNPSINEGTDFNTMKIFSDFKLASIGNE-----DNM 444
 631 IDFIISGQKKYLEL-QRIRDNISYNNLFYTT-----EDLSLNNVAIGGIPAKKYLEHG 684
 445 MFMIKITNYLVKGFAPDVRSTINLSGP----- 471
 685 LF-----SEYRQDGTIPVYVSTLNSGPDWIMQMKYKSLGRIGEVHIKONKLSVDNVL 740
 472 GVTGAYQDLMF---KONSTNIHLEPE-----LRNPEFPKTKISQLT---EQEITS 518
 741 GYVASSNKDKSNFNLNPFVSGINDITPDESSAVARNNDINKILFEKINCHVPEKLPTS 800
 519 LMSFNQARAKSOFBEYKKGYPFEGALGEDNDLFAQTVLDKDYVSKKILSSMKTENKEYI 578
 801 LY-----YEIDRSRFFQGW---DNKSIKVTWINKDLI--KQINLLTSSNDV 844
 579 HVIYVQLQGDKISVEASCNLFSDPYSSILYQKNIEGSETAYYYVADAEIKEIDKYRIPY 638
 845 KLLIKL--DRELYAISSKI-----DNPLASRISRTLQ-QLANYVTSNTFEPENTINFIY 896
 639 QISNKN-----IKLTFIGHGKSEFNTDTPANLDVS-----LSSEIET 677
 897 DFYRKQDQLLSAIDL-----FSRNDADTKIIVVYNSVMKXVFLREVISC 942
 678 ILMKADISPKYIEINILGCMFYSISAEITYGKLLTKDKRVSELMPISODSI-- 735
 943 VLRSKKVD---SYINEN-----KQNSKEDA--GAURDYAKLKKKELFSLMDDDDGYKK 990
 736 TVSANOVEYRINEEGKREILDHSGKWINKEESIIKDISKEYISFNP-----K 783
 991 IITNAY-----IKERDKL-----SGIYIENSLIGSHESFDIIRSNQHEWGLSTVEQFK 1042
 784 ENKIIKVKSYLHLSLTLQEIIRNNANSDDIDLEKKVMLTECINVAS-----NIDRQIYE 838
 1043 KFEFYKSE-LSSAKSIFDDIKNK-YITDPETKRVLYHQLDSDIKERIAFLDISHYAYP 1100
 839 GRIEEAKNLTSDSINYKNEFKLIESI-----SDSLYDLKHQNGL 878
 1101 GSLEKLQLS---GYVFSIDNIIAEYLLASYGVSHGVYVYVAPSPDKLLELRHTK 1156
 879 DSHFISFEDISKTEGPRIRFINKETGSI---FIETEKEIFSEYATHISKEISNIKT 935
 1157 SNSEWI--EKITP-----YVVDILSDNVNVLRPPLSEEQKILNDIKLEISKVS---EQ 1207
 936 IFDNVNGKLVKYNLDAAEVNTLNGAFFIQSILIEYNTTK-----ESLSNLSV----- 983
 1208 YFMKLTQKSSVIGIKYVDFDYRNLFLSLPILNQNLTLPFMYRYFEMLYDIHIGIEN 1267
 984 -AMKVQVYQLFSTGLNTITDASKVVEL 1010
 1268 KANREFIYKSFSSLNLDLFLINDERVNL 1295

RESULT 7
 AAR07503
 ID AAR07503 standard; protein; 1254 AA.
 XX
 AC AAR07503;
 XX
 DT 25-MAR-2003 (revised)
 DT 06-FEB-1991 (first entry)
 XX
 DT Merozite apical-end-localised protein (MABP) insert 5.3.
 DE
 XX Malaria; vaccine; Duffy blood group.
 XX
 OS Plasmodium vivax.

PN W09011772-A.
XX
PD 18-OCT-1990.
XX
XX 05-APR-1989; 89US-00334041.
PF
XX 05-APR-1989; 89US-00334041.
PR
XX 06-APR-1989; 89US-00334270.
XX
PA (UYNV) UNIV NEW YORK STATE.
XX
XX Barnwell JW, Galinski MR, Wertheimer SP;
PI
XX WPI; 1990-334616/44.
XX N-PSDB; AAQ06328.
XX
XX Malarial apical end merozoite proteins and peptide(s) - used for
PT developing cpds. for treating, preventing and diagnosing malarial
PT infection.
XX
XX Claim 2; Fig 1a; 66pp; English.
XX
XX A MAEP compound having a binding affinity for a Duffy blood group antigen
CC of primate red blood cells, is antigenic for the complete protein, and
CC may be used in diagnosis, treatment and vaccination against invasion by
CC P.vivax and P.knowlesi. (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 1254 AA;
Query Match 5.7%; Score 294; DB 2; Length 1254;
Best Local Similarity 19.6%; Pred. No. 2.1e-07;
Matches 213; Conservative 198; Mismatches 408; Indels 269; Gaps 48;
QY 64 KSGRNKALKKFEYLTMEVLEKNNSLTPVEKNLHFIWIGGOINDTAINYNQKDVNS 123
DB 14 KKSIEKAYEKMGN--TLKELEKMD--EKN-----IEKEVEBAQIQY----- 52
QY 124 DYTQVYDQNAFLINTLTKTIVESATNTLESFRENLANPDPDYKFKYKMEIYDQK 183
DB 53 ---KRIFIDHV-----NLMNDEVEKSVINKEIE-LYKKE 84
QY 184 KHFIDYKSOIEENPEFIIDNIKTLSNEYSKDLBALNKYIBESLN-----KITANNNG 238
DB 85 ---IDEIKQKNEVQDGTNF---YTEQNSATQSKAK-IEQFINIATTKGISTDQ 137
QY 239 DIRNLEKFADE-----DLVLYNQELVE-RWNLAASDILRISMLKEDGVDVLDILPG 292
DB 138 DINELESIKEEVHKNLQVQESNSMEERKQILSMKDLILN----- 180
QY 293 IQPDLFKSINKPDSITNTSWEMIKLEAIMKYEYIPGYTSKNFMDLDEEVORSFESALSS 352
DB 181 -----NSETTAKESINTQNALGFRENATK-----LNKTDLLQRYAAMEEA 224
QY 353 KSDKSEIFLPLDDIKVSPLEVKIAFANNVINQA-----LISIKD---SYCSDLVINQIK 404
DB 225 KAHKNNDIALEDAQIDTEVSKIEQINREIMNKKDEIKSYLSEIKYOKCTTEISNKR 284
QY 405 NRYXI-----LNDNLNPSINEGTFD-----NTTMKIFSDKLA 436
DB 285 GKQKIEFLEKFPNEBSNSKNVINEINIRSEQYLKJEDAEKCASTKVELFKHET 344
QY 437 SISN--EDNMMFMIKITNLYKVGAPDVRSTINLSGPGVYT--GAYODLMPKDNSTNIH 492
DB 345 TISNIFRESEILGETVSKQKINKAEDIMKEIERHNSIEIQTVKGFQENLKNLNEPHNYD 404
QY 493 LLEPELRNFPKPKTKISOLTEQBITSLWSFNQAPAKSQFEEYKGYEGALGE---DDNL 549
DB 405 NAEDELANDKSTNAKV--LIETNLESV-----KHLSEITNIKQ-----GEKIYSKAK 451
QY 550 DPAQNTVLDDKVSKILSSMKTKENKEVYHVIQDQKISYEASCNLPKSDPVSSY-LY 608
DB 452 DIMQKIKATSENTAKTLEKVDQDSNYVNYL-----NQITERNLIVTEKRLNGIDST 506

QY 609 QKNEIGS--ETAYVYVADAE-IKEIDKYRIPYQISNKRNIKLTFIGHGKSEFNT----- 660
DB 507 ITNIEGALKESKNGYEIGFLEKLEEGKRNKRLKVDITKKSINST-VGNFSLFNFDLQ 565
QY 661 -DTPANLD-----VDSLSSEIETILNLAKADI-----SPKYI-----EINLL 696
DB 566 YDFNKNINDYENKMGGEIYNFEFGLSNKISENLRNASENTSDYNSAKTLRLLEAQEKVNL 625
QY 697 G-----CNMFYSISAEETYPGKLLKIKORVSELMPSISQDSITVSANQYEVRAINBEGR 752
DB 626 NKEEANKYLRDVKKVESF--RPIFNMKESLDKINEMIKKEQLTVNEGHGNVQLVLENIK 683
QY 753 BILDHSGKWNKEESIIKDISKEYISFNPKENKIIVKSKYLHELSTLLQELIRNANSSD 812
DB 684 ELVDE-----NNLSDIILKQATGKN-----BEIQKITHSTLKNKAKTILGHVDTSAKYVG 732
QY 813 IDLEKMYLDE---CEINVASNIDROIVEGRIEEAKNLTSDSINYIKNEFKLIESISD- 867
DB 733 IKITPELALTELLGDALKTAQELKFSKNVULETENMSKNT-----NELDVHKNIQDA 787
QY 868 ---SLYDLKHQGLDSDSHFISFEDISK-TENG-----FRIRPINKETGNSIFITEKEIFS 919
DB 788 YKVALEILAHSDIEDTKQ---KDSKLIEMGNQIYLVKVLINGYKKNKISSIKSEKSAVS 843
QY 920 EYATHISK---EISNI---KOTIFPNVNGKLVKKNVLDAAHEVNTLNSAFIQ----- 966
DB 844 VKIGNVSKHSELKSKITCSDKSYDNI-----IALEKQTELOQLNRSFTQKETTNTNSDS 896
QY 967 SLIEWNTKESLSNLSVAMK-----VQVYALFSTGLTNTITDASKVVVELV 1011
DB 897 KLEKIKTDFESLKNALKTLEGEVNALKASSDNHEHVOSKEPVPNALUSEIEKEETDIDSL 956
QY 1012 STALDETI 1019
DB 957 NTRALDELL 964
RESULT 8
AAW24575
ID AAW24575 standard; protein; 1254 AA.
XX
AC AAW24575;
XX
DT 25-MAR-2003 (revised)
DT 10-NOV-1997 (first entry)
XX
XX Merozoite apical-end protein clone 5.3.
XX
XX Merozoite apical-end protein; MAEP; Plasmodium vivax; antigen; malaria;
KW human; erythrocyte; antimalarial vaccine; antibody; blood stage parasite;
KW Duffy blood group antigen; red blood cell; therapy.
XX
OS Plasmodium vivax.
XX
XX US5646247-A.
XX
PD 08-JUL-1997.
XX
XX 04-OCT-1991; 91US-00792865.
XX
XX 05-APR-1989; 89US-00334041.
PR 06-APR-1989; 89US-00334270.
PR 03-APR-1990; 90WO-US001849.
PR 02-NOV-1990; 90US-00608639.
XX
XX (UYNV) UNIV NEW YORK STATE.
XX
XX Galinski MR, Barnwell JW;
PI
XX WPI; 1997-362995/33.
DR N-PSDB; AAT80072.
XX
XX Plasmodium merozoite apical end protein - useful as antigen for
PT

PT	production of anti-malarial vaccines.	
XX	Claim 1; Col 29-38; 68pp; English.	
PS	AAW24575 and AAW24576 represent the merozoite apical end proteins (MAEP)	609 QXNIEGS--ETAYYYVADAE-IKEIDKYRIPYQISNKRNIKLTIFIGHGKSEPT----- 660
XX	isolated from two different Plasmodium vivax strains. These proteins are	609 QXNIEGS--ETAYYYVADAE-IKEIDKYRIPYQISNKRNIKLTIFIGHGKSEPT----- 660
CC	the antigens of the invention, and immunoreact with antibodies against a	507 ITNIEGALKESGNGVEIGFLEKLEIGKVRKLVDTIKTSINST-VGNFSLFNFDLQ 565
CC	native MAEP sequence. P. vivax is one of the four malarial species that	507 ITNIEGALKESGNGVEIGFLEKLEIGKVRKLVDTIKTSINST-VGNFSLFNFDLQ 565
CC	infects humans, and is difficult to target for a vaccine, as it cannot be	661 -DTFANLD-----VDSLSEIETITINLAKADI-----SPKYL-----EINLL 696
CC	cultured in vitro. The preinvasion orientation of malarial merozoites	661 -DTFANLD-----VDSLSEIETITINLAKADI-----SPKYL-----EINLL 696
CC	indicates that the apical end plays an important role in the invasion	566 YDFKNKIINDYENVMGEIINEFEGSLNKISENLNASENTSDYNKAKTILRLAQEKYNLL 625
CC	process. The MAEP protein binds to the surface of susceptible	566 YDFKNKIINDYENVMGEIINEFEGSLNKISENLNASENTSDYNKAKTILRLAQEKYNLL 625
CC	erythrocytes from P. vivax susceptible humans and primates, and also	697 G----CNMFYSISAEETYPQGLLLKIKDRVSELMPSISQDSITVSANQYEVRIEENK 752
CC	binds to rabbit erythrocytes. The antigen can be used for the production	697 G----CNMFYSISAEETYPQGLLLKIKDRVSELMPSISQDSITVSANQYEVRIEENK 752
CC	of antimalarial vaccines. The antigens are involved in the invasion	626 NKEEBANKYLVRVKKVESF--RFIFNMKESLDKINEMIKKQLTVEGHGNNVQLVENIK 683
CC	process, and are immunohistochemically reactive with antibodies raised against	626 NKEEBANKYLVRVKKVESF--RFIFNMKESLDKINEMIKKQLTVEGHGNNVQLVENIK 683
CC	malaria (particularly P. vivax) blood stage parasites. Synthetic	753 EILDHSGKWKNEBIIKIDISKEVINSNPENKLIKVKYLHELSTILLOIRNANSSD 812
CC	proteins, polypeptides, peptide fragments and analogues of these antigens	753 EILDHSGKWKNEBIIKIDISKEVINSNPENKLIKVKYLHELSTILLOIRNANSSD 812
CC	can be used similarly. As the antigens specifically bind to a Duffy blood	684 ELVDE-----NNLSILKQATGKN-----BEIQKITHSTILKNKAKTILGHVDTSAKTVG 732
CC	group antigen (the antigen present on the surface of susceptible	684 ELVDE-----NNLSILKQATGKN-----BEIQKITHSTILKNKAKTILGHVDTSAKTVG 732
CC	mammalian red blood cells), and are necessary in the process of invasion	813 IDEKKWMLTE-----CEINVASNIDRQIVEGRIEAKNLTSDSINYIKNEFKLIESID- 867
CC	of red blood cells by merozoites, they can be used to inhibit the	813 IDEKKWMLTE-----CEINVASNIDRQIVEGRIEAKNLTSDSINYIKNEFKLIESID- 867
CC	invasion of red blood cells by a malarial organism. The antigens can also	733 IKITPELATTELLGDAKLTQAFELFESKNNVLENNKNT-----NELDVHKNIQDA 787
CC	be used in a method for inhibiting invasion of susceptible mammalian	733 IKITPELATTELLGDAKLTQAFELFESKNNVLENNKNT-----NELDVHKNIQDA 787
CC	blood cells by malarial merozoites, and in a method for inhibiting the	868 ---SLYDLKHQGLDSDSHFISPEDISK-TENG-----PRIFINKETGNSIETETEKEIFS 919
CC	propagation of a malarial organism in susceptible red blood cells.	868 ---SLYDLKHQGLDSDSHFISPEDISK-TENG-----PRIFINKETGNSIETETEKEIFS 919
CC	(Updated on 25-MAR-2003 to correct PF field.)	788 YKVALEILAHSDIETKQ-----KDSKSLIEMGNQIYLVKVLINQYKNKISSIKSEEAVS 843
XX	Sequence 1254 AA;	920 EYATHISK---EISNI---KDTIFDNVNGKLVKKNLDAHEVNTLNLSAPFIQ----- 966
QY	Query Match 5.7%; Score 294; DB 2; Length 1254;	844 VKIGNVSKHSELSKITGSDKSYDNI-----IALEKQTELQNLNRSFTQKTNNTSDS 896
Best Local Similarity 19.6%; Pred. No. 2.1e-07;		967 SLIEVNTYKESLSNLSVAMK-----VOVYQALFSTGLNTITDASKVVELV 1011
Matches 213; Conservative 198; Mismatches 408; Indels 269; Gaps 48;		967 SLIEVNTYKESLSNLSVAMK-----VOVYQALFSTGLNTITDASKVVELV 1011
QY 64 KSGRNKALKPKPEVLTNEVLELKNLSLTPVEKNLHFWIGQINDTAINVQWKDVNS 123		897 KLEKINTDFESLKNALKTLEGEVNAKASSDNHEHVQSKSEVPNPALSEKEKETDIDSL 956
14 KKSIEKAVEKMGN--TLKELEKMD--EKN-----IEKEVEEAQIQY----- 52		1012 STALDETI 1019
124 DTVYKVFVDSNAFLNTLTKKTVESATNTNLTESFRENLDPEFDYNKFKRMELIYDKQ 183		957 NTALDELL 964
53 ---KRFIDHDV-----NLMNDEVERKSKIVMEKIB-LYKKE 84		
184 KHFIDYKSOEENPEFIIDNIITKYLNGEYKOLEALNKVIERSLN-----KITANNGN 238		
85 ---IDEIFKQNEYKQGTSPN---YYTEQNSATQSKAK-IEQFINIATTKGTSTSQ 137		
239 DIRNLEKFADE-----DLVRLYNQELVE-RWNLAASDILRISMLKEDGGVYLDVILPG 292		
138 DINELESIKEEYHVKLQLVKQESNVEEMRKQILSMKOLLIN----- 180		
293 IQPDLFKSINKPDSITNTSWEMIKLEAMKYKEYIPGVTXKFNMDLDEEVORSEFESALSS 352		
181 -----NSETIAKEISNNTQNALGFRENAKTY-----LNKTDLELQORVAAMIEEA 224		
353 KSDKSEIFLPDDIKVSPLEVKVIAFANNVINOQ-----LISLKD--SYGSDLVINQIK 404		
225 KAHKNNDIALEDAQIDTEVSKIEQINREIMNKDEIKSYLSEIKYKDKCTTISNSKR 284		
405 NRYKI-----LNDNLNPSINEGTDF-----NTTWKIFPSOKLA 436		
285 GDKLEFLEKFKPNEBESNKNVINEINENIRNSEQYLKDIEDAKQASTKVELFKHET 344		
437 SISR--EDNMFMKITYNLKVGFPADVRSTINLSGPGVYT--GAYQDLMLFKONSTNIH 492		
345 TISNIFKESEILGVETKSKINKAEDIMKEIERHNSBIQTQVKGQENLKNLNEPHNYD 404		
493 LLEPELRNEFFKTKISQITQETISLWSFNQARAKSQFEEYKGYFSGALGE---DDNL 549		
405 NAEDELNDKSTNAKY--LIETNLESV-----KXNLSEITNIKQ-----GEKITYSKAK 451		
550 DPAQNTVLQDYVSKKILSKMTNRKEYITHYIVQLQGDKISYEASCNLFSKDPYSSI-LY 608		
452 DIMQKIKATSENTAETLEKVKDDQSNVNYL-----NOITERNLIVTERKRLNGIDST 506		

RESULT 9	
ABU48941	
ID ABU48941 standard; protein; 4688 AA.	
XX AC ABU48941;	
XX AC	
DT 19-JUN-2003 (first entry)	
XX DE Protein encoded by Prokaryotic essential gene #34468.	
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.	
XX OS Ureaplasma urealyticum.	
XX PN WO200277183-A2.	
XX PD 03-OCT-2002.	
XX PF 21-MAR-2002; 2002WO-US009107.	
XX PR 21-MAR-2001; 2001US-00815242.	
XX PR 06-SEP-2001; 2001US-00948993.	
XX PR 08-OCT-2001; 2001US-0342923P.	
XX PR 08-FEB-2002; 2002US-00072851.	
XX PR 06-MAR-2002; 2002US-0362899P.	
XX PA (ELIT-) ELITRA PHARM INC.	
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;	
XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;	
XX DR WFI; 2003-029926/02.	
XX DR N-PSDB; ACA52811.	
XX PT New antisense nucleic acids, useful for identifying proteins or screening	

PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 76865; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: the sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Seq Sequence 4688 AA;

Query Match 5.6%; Score 290.5; DB 6; Length 4688;
 Best Local Similarity 18.8%; Pred. No. 1.8e-06;
 Matches 252; Conservative 220; Mismatches 433; Indels 437; Gaps 65;
 QY 13 VYKFRIOEDYVAINLALE-EVHNMBSSVVEKYLKIDINLTNYLNT---YKSGR 68
 Db 403 LYEVIKKTNEVDVNLKQIPYHRS-----INNLSNALNTPYQYTKGD 449
 QY 69 NKALKKFEKLTWE---VLELKNSS---LTPVEKNLHFIWIGGOINDTAINYNQWKD 120
 Db 450 INLIKAVPYVYNQVYGFQDQNOEHQILAKVKKDGTAFDTGALNNSNYSLDKIVS 509
 QY 121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTLESFR-----ENL----- 161
 Db 510 VSNPQNLV---SNFDLTSQKQKLIKPAANASVDTSTKTOILENLNLDLQKLVATFV 566
 QY 162 --NDPEPDYKFKRMEIYDKQ---KHFIDYKSOIBEEN----- 197
 Db 567 DNNDKYYKVVANDQNNKIIFSDNLPKGYI-YHLAKVENNDLNKVINLNDPDKIID 625
 QY 198 -----PEFIIDN---IKTVLSNYSKDL---EALNKY-----IEESLN 230
 Db 626 KRDLNILDSPDFTYDNDGNLEIHTQLANDLNDLQKALNNANVKGIIVDQDGIHEID 685
 QY 231 KITANNNDIRNLEKFADELVR--LYN-CELYERNWLAASDILRISMLKEPDGGVYLDV 287
 Db 686 VSDANGKVIPTKLANNDPTKPIYTLKKVYKQNNQPNIDLISEEQSGDNHISFK- 744
 QY 288 DILPGIQDPLFKSINKPD-----SITNTSMEWIKLEALMKYKEYIPGYTSKGFMDLDEEVQ 343
 Db 745 -----KPTTAKTKENDDYVEISFNPSLANKIKLTKFTDD-----NNTWTKTVEAS 791
 QY 344 RSPESALSKSKSEIFLP-----LDDIKV-----SPL----- 371

Db 792 IGLDGKAIKFTSDDAIFAPDHPKVTLTKEADNKKVANIDEISPLDRIVNKKQNGNWNADN 851
 QY 372 --EVKIAFANNYSI-----NQALISLKDSYOSDLVINQIKRY-----KI-- 409
 Db 852 KHEFKIPDQKNDLTAVYKDKNNNEIHVPKIDDDKGVVWPNNNLFDPNKIYDFDKVD 911
 QY 410 LNDNLNPSINEGDFNTWKIFSD-----KLASISN-EDNMMPKIKITNVLKVG 457
 Db 912 LNEYPNKILTDRNSINKDVSAINDGVQDQARKLVKAPAVSNVTVAINFQVNLFDNLKUS 971
 QY 458 ----RAPDVRSTNLSPGCVYTGAQ-----DLMFKDN 487
 Db 972 YNQFALTUAKVNDLNDTQKIATYDPKTNVYKLNDFDTHLDANTKYVVDVLELNNK 1031
 QY 488 STNHLLEPELRNFEPPKTKISOLTEQEITSLSWFSNQARAKSQ-----FEYKGYFEG 541
 Db 1032 EKPIKLIKDVLPNETTSATTIN---PPIWTKEDVVVKTNNNDITFIETDK--DN 1084
 QY 542 ALGEDDNLDPQANTVLDKDYVSKILLSSMKTRKNEYIHYIVQOGDKISYEASCNLFSDK 601
 Db 1085 ILKNDQKI-YAQLALMDDDLDDTDVINPLVNFSEKT-----NKIASINGLDLKGNS 1133
 QY 602 PYSILYQKNIESETAYYYVADABLK-----EIDKYRIPYQISNKRNIKLTFIGHGK 655
 Db 1134 KYST-----KNL-----YFLNDQNKQVYLFKXNDVTKYB-QHFTNPHKINJSF--NK 1177
 QY 656 SEFWTDF---ANL-----DVSLSSEIETILMLAKADISPKY 690
 Db 1178 SAVEQDIFAMHANLFDYKDYDOKLRINEDVKIYYQINIDTKNELO--IGYKV-VANNK 1234
 QY 691 IEINLLGCMFSYSISABETYPGKLLKIKDRVSELMPS-----ISQDSITV 737
 Db 1235 IKFNLVG-----LKEKTIYVIKKLEALNKSASSIVNSEFDLLDPNFTSNKNTLV 1287
 QY 738 SANDY-----EVRIIEGKREILDHSGKWI-NKEESIIKDISKEY 777
 Db 1288 GLNSIDNNNDQTEPIINAKINIGDDFQDNOQVLLIYVSNNDNKEIKSAVTLIKQRYQF 1347
 QY 778 ISFPNPKNLIIVKSY-----LHELSTLLQEI-----ENNA-----NSSDIDLEKKVML 821
 Db 1348 EFSNLKLNRYTFKSIYVETNNQTLHKLDLTLTQFSINPNNNAVSLKKNNTNIEITKILV 1407
 QY 822 TECEINVASINDROIVEGRIE--EAKNL--TSDSINYIKNEFKLIESISDLYDLKHQNG 877
 Db 1408 -----NDQGLISAKIEVDIDVNLTAUKPNIVYO-----LENSDANNLKLATN 1454
 QY 878 L-----DSSHFIISPEDI-----SKTENGFRIRPINKETONSIF----- 910
 Db 1455 AQVVVENNKKFLKFDLAVSLKINQNVYIKESIFSNKPTNAY-FNPTNNKTNIVSYDEQN 1513
 QY 911 -----IETEKEIF-SEYATHISKEISLNK-----DT 935
 Db 1514 KISLSNNINPTSYKPTINKNDKSVNIDVDLQVDKQLLANQYLRKLKQLNDNKTVMWTD 1573
 QY 936 IFDVGNGKLVKKNLDAAEVNTLNSAFFQSLI---EYNTTKESLSN-LSVAMKVQVYA 991
 Db 1574 ILFNNAKISPKLS-NLIH-----NRAYELEGLYFDQNSVNDMTNQNISFNSKIHKPK 1627
 QY 992 QLFSTGLNTIT-DASKVVELVS 1012
 Db 1628 IEFPSLTTINDTNNAIKTVS 1649
 RESULT 10
 AAB18171
 ID AAB18171 standard; protein; 1979 AA.
 XX
 AC AAB18171;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:28.
 XX

RESULT 12
 ABU24510
 ID ABU24510 standard; protein; 1178 AA.
 AC ABU24510;
 XX
 XX 19-JUN-2003 (first entry)
 XX
 XX Protein encoded by Prokaryotic essential gene #10037.
 XX
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 XX Clostridium botulinum.
 XX
 XX WO200277183-A2.
 XX
 XX 03-OCT-2002.
 XX
 XX 21-MAR-2002; 2002WO-US009107.
 XX
 XX 21-MAR-2001; 2001US-00815242.
 XX
 XX 06-SEP-2001; 2001US-00948993.
 XX
 XX 25-OCT-2001; 2001US-0342923P.
 XX
 XX 08-FEB-2002; 2002US-00072851.
 XX
 XX 06-MAR-2002; 2002US-03628999.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX Wall D, Irawick JB, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 XX WPI; 2003-029926/02.
 XX
 XX N-PSDB; ACA28380.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 XX for homologous nucleic acids, required for cellular proliferation to
 XX isolate candidate molecules for rational drug discovery programs.
 XX
 XX Claim 25; SEQ ID NO 52434; 1766pp; English.
 XX
 XX The invention relates to an isolated nucleic acid comprising any one of
 XX the 623 antisense sequences given in the specification where expression
 XX of the nucleic acid inhibits proliferation of a cell. Also included are:
 XX (1) a vector comprising a promoter operably linked to the nucleic acid
 XX encoding a polypeptide whose expression is inhibited by the antisense
 XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
 XX polypeptide or its fragment whose expression is inhibited by the
 XX antisense nucleic acid; (4) an antibody capable of specifically binding
 XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 XX proliferation or the activity of a gene in an operon required for
 XX proliferation; (7) identifying a compound that influences the activity of
 XX the gene product or that has an activity against a biological pathway
 XX required for proliferation, or that inhibits cellular proliferation; (8)
 XX identifying a gene required for cellular proliferation or the biological
 XX pathway in which a proliferation-required gene or its gene product lies
 XX on a gene on which the test compound that inhibits proliferation of an
 XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
 XX compound's activity; (11) a culture comprising strains in which the gene
 XX product is overexpressed or underexpressed; (12) determining the extent
 XX to which each of the strains is present in a culture or collection of
 XX strains; or (13) identifying the target of a compound that inhibits the
 XX proliferation of an organism. The antisense nucleic acids are useful for
 XX identifying proteins or screening for homologous nucleic acids required
 XX for cellular proliferation to isolate candidate molecules for rational
 XX drug discovery programs, or for screening homologous nucleic acids
 XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 XX the target prokaryotic essential genes. Note: The sequence data for this
 XX patent did not form part of the printed specification, but was obtained
 XX in electronic format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 1178 AA;
 XX SQ

QY 918 -----FSEYATHISKEISNIKOTIFDNVNGKLVKVNLDAAHEVNTLNLSAFFIOSLIEYNT 973
 Db 1046 KDITGRVALELD---CNGNFTMRDDFNGGIRPTNLSGGET-----FLTSLA----- 1091
 QY 974 TKESLSNLSVAMKQVQYQA-----LFTGLNTITDASKWELVSTALD 1016
 Db 1092 -----LALALSSQILKGSPLFFFLDEGFGTLD--SELLDTVNTSLE 1133

RESULT 13

ID ABU25521
 ABU25521 standard; protein; 1182 AA.

XX AC ABU25521;
 XX DT 19-JUN-2003 (first entry)
 XX DE Protein encoded by Prokaryotic essential gene #11048.
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX OS Clostridium difficile.
 XX PN WO200277183-A2.
 XX PD 03-OCT-2002.
 XX PF 21-MAR-2002; 2002WO-US009107.
 XX PR 21-MAR-2001; 2001US-00815242.
 XX PR 06-SEP-2001; 2001US-00948993.
 XX PR 28-OCT-2001; 2001US-0342923P.
 XX PR 08-FEB-2002; 2002US-00072851.
 XX PR 06-MAR-2002; 2002US-0362699P.
 XX PA (BLIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA29391.

PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 53445; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation. (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 1182 AA;

Query Match 5.3%; Score 276.5; DB 6; Length 1182;
 Best Local Similarity 20.7%; Pred. No. 1.7e-06;
 Matches 222; Conservative 205; Mismatches 372; Indels 275; Gaps 55;

QY 8 QLOQWYVVFRIQDEYVAILNALBEYHNHMSSESVVEKYLKDINNLTDNYNTYKSG 67
 Db 199 QLOQLKOLKSKL--SQVDGITE--EYNNINQELIKNLE-KDKNALDLAQSYEES- 252
 QY 68 RNKALKKFKFVLTMEVLELKNNSL---TPVENLAFIMIGGOINDTAINYNQMDVNS 123
 Db 253 --KTV--YESQLKLEKRELKNELDLKNSIEKIALV-----ENDSNKKVDPY--IS 301
 QY 124 DYTQVYFDSNAFLINTLTKTIV-----ESATNLTLESFRENLD--PFDYNKFKYKRM 176
 Db 302 VONLEKKIDEDSPAVSNLEKKLAILNQLELVTKRYEKINKIKNEVPKLSSEKI-RLQQ 360
 QY 177 BIIVDKOKHFDYKSOIENPEFIDNIIKTVLSNEYSKOL-----EALNKVIESLNK 231
 Db 361 AIKLEBELVLLRELKELKESG---INLNKTKVELEKVKQVSESCKDAVTKSKEVEGK 416
 QY 232 ITANNGNDIRNLKFADEDLVLYNQELVER--WNLAAADILLRISMLKDDGGVYLDVIL 290
 Db 417 IDKVNISAEILKQKIFLAYEYEDKDYNNVLEBNOKLNKLEBI-----LKTENINLKV--- 468
 QY 291 PGIQDFLKSINKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKFNFDLDEEVQSFESAL 350
 Db 469 -----RYTDKNDVNNLENLSLHLDVLKK-CPG--KSAULL---LKSEVVTTEL 513
 QY 351 SSKSDKSEIFLPLDDIKVSPLEKVIAPANNVINOALISLSDYSCDLVINQIKNYKIL 410
 Db 514 RNK-----ANNTKENE-----IKKS-----SIQDELKII 537
 QY 411 NDNLNPSINEGTQFNT--TWKIFSDKLA-SISNEDNMFMKITYNKLVKGFADPVRSTIN 467
 Db 538 LE-----SKPNTREVNLLNKLNNRKNRDOLEKELEELKYNL--ASELRREUK 586
 QY 468 LSGFQVYVYGA-----YQDLLMFKDNSTNIHLLELRLNFFPPFKTISQLTEQ---EITSLW 520
 Db 587 ENMPCPVCGSKHENHDITKYDE--NISFVKEKLEKEKISIRNTEELNAKVSQYL 643
 QY 521 SFNQARAKSQPEYKGYFEGALGEDDNLDFQNTVLDKDYVSKTILSMKTRNKYIHY 580
 Db 644 SIEMKTK-BLEDVK-----GKLGE-----IPSSQLLKLDEBQRKALLKSNIQE--- 688
 QY 581 IVQLQGDKI SYEASCNLFSDPYSSILYQKNIEGSTATYVYVADAEIKEDKRYPIQI 640
 Db 689 -----WKEKESTENK--VTIAKEEKNIKEELKIRE 719
 QY 641 SNKENIKLTFIGHKSEFNTDTFANLDVPSLSSEIETILNLAKADISPKYIEINLCNM 700
 Db 720 SLNNYKLT-----RDNLNIEGLEKCK-KIKQVGLGKTI---- 755
 QY 701 PSYISIAEETYPKLLLIK-----DRVSELMP-----ISODSITVSANQVETINEE 749
 Db 756 -----TKVNNLSKVKRENEKELELLSTGSHNLLKNRDSLDIKIRENESKLHE- 805
 QY 750 GKREILDHSGKWNKEESIIKDISKEYISFNPKENKIIVKSKYHELSTLQELRNAN 809
 Db 806 --IEIEIKARELYVVEKKLSRONKYKEVISITKGD-----LAKNLLHNVENICKILEQEE 859
 QY 810 SSDIDLEKKVMTCEINVASNIDROIQVEGRIEEAKNLTSD---SINYI--KNEFKLIES 864
 Db 860 SSKKLEBEQL--EYKKNIA---EKNIDGRUKTAQEQKQDKQSTINKLUAENKFSIYA 914

QY 865 ISDSLDP-----LKHONGLDSDHSPISDISKTB--NGFRIRFINKETGNSIFI 911
DB 915 VRALLDTPDAVKKLIBELERE-----EQKULSPFIKSKSKELNG---RSIKKZ-----YFE 964
QY 912 ETEKEIFSEVATHISKEISNIKDTIPDNVNGKLVKYNLDAAEHVNTLNSAFTIQSLIEY 971
DB 965 QKDEIY-----NLKVEIGNISKDIGANQNLITLKDSLD---KINDFNKQL---KVVEH 1013
QY 972 NTKTESLNLVAMKVQVYALFSTG-----LMTITDASKVVLVST 1013
DB 1014 NV--DLLEELDKVQGNKEFVYSTNQIKYIALEBASKRLDGTIKGRYALRIDST 1065

RESULT 14
ABU24531
ID ABU24531 standard; protein; 1279 AA.
AC ABU24531;
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #10058.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Clostridium botulinum.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-029926/02.
XX N-PSDB; AC28401.

XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 52455; 1765pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX

Sequence 1279 AA;
Query Match 5.3%; Score 275.5; DB 6; Length 1279;
Best Local Similarity 21.2%; Pred. NO. 2.2e-06;
Matches 239; Conservative 187; Mismatches 382; Indels 317; Gaps 62;
QY 25 VAINALBEEYHN-MGESSVVEKYKLDKNNLNTYNTYKSGRNKALKKFK--EYJTM 81
DB 310 VELKNILDSYENEFIKKGFIFLYPSMKALSNLVILDKKYEAKKRRDLIDFNDIEHLCL 369
QY 82 EYLEKNNSLTPEVKNLHFIWIGGOI--NDTAINTINOWKOVNSDYTVKVFYDSNAFLIN 139
DB 370 SILTDKNSE-----GHIIPSDIALDYRKFAEV-----LIDEYQDSN--LVQ 409
QY 140 TLKKTIIVE-----SATNNTLESFRE--NLNDPEFY---NKF-----YKRMET 178
DB 410 EVTMSVSRVKGYSFYNGQL-MFNEBEINLEPQICLDIPRMVGVGVKQSIYFRQA- 467
QY 179 IYDKOKHFDIYKSOIEE-----NPEFIIDNIITKYSNE--YSKOLEALN 222
DB 468 ---KPEIFLDKNEYSEEGTKNRKVLKPKNSRKEVINGV--NYLPKQINSKITGELD 522
QY 223 KYIEESLNKITANNNDIRN-----LEKPAD--EDLVRLYN---QELVERNL---AA 267
DB 523 -YTEEEALVGSAGYGEVKGEPLELCLMDKKYIESEEVLEKYNVDEEALDNIQLEGLV 581
QY 268 ASDILRISMLKEDGV-----YLDVILFGIQDPLFKSINKPDSITNTSWE 313
DB 582 AKKIQLGVNLGGLGVDFKGLGEYRNLYQYRDIVL-----MRATSNWA 626
QY 314 MIKLEAIMKYKEVPGYTSKNFMDLDEEVQSFESALSSKSDKSEIFLPLDDIK-VSPLE 372
DB 627 PIFVEELA--KEGIPVFADTNGYFTAIKTIWSLLOI-----IDNPLQIPLLSVLVR 678
QY 373 VKIA-FANNSVINQALISLSDY--CSDLVINQIKNR-----YKILNDN 413
DB 679 SPIASTDDELDIRMVKNKITFECEWELIYRLYKNEKLDYVSFYIEDENKINKIKDM 738
QY 414 LNPSINEGTDFTNTMKIFSDKLASISNEDNMMPMKITNYLVKVPADPVRSTINLSGPV 473
DB 739 NEKLKNKICSFIEKLKWEKSIHIDIDFIFWFL-----YVETGYG----- 780
QY 474 YTGAYQDLMPKDNSTNIHLBEPENFPFKTKISQLEBTSLSWFSNQARAKSQFEE 933
DB 781 YAGALQ---AGEORQANLRII-----FORAK--QYAKTSYKGLFNFINFKLKFS 827
QY 534 YKGYEGALGEDNDLDFQAQNTVLDKDYVSKKLSLMSKTRNKKEYIHIVQLQDQKISYA 939
DB 928 GDMGSAK-IUGENENV-----VRIMSHKSKGLEF--PVVILSG-----T 864
QY 594 SCNLFSKDPYSSILYOKNIEGSETAYYYVADAE-----IKEI-----DKY 634
DB 865 GKFNMTDLANKNLFHRDL-----GYGPDVDPERRIAYPSLVKNIKNKIRLETSEEM 919
QY 635 RIPPYISNKNENIKLTFIGHCKSEFNFTDTANLDDVDSLSSEIEITILNKAADISPKYIEIN 694
DB 920 RILYVALTRAREKLIITGL-----INNMDKTVDWNLNLS-----DKN 957
QY 695 LLGCNMFYSISAEETY---PGKLLKIKDRVSELMPSISQDSITVSANQYEVRIINEEGK 751

958 ----KVPEYAVMSGKTYLDWIGPALKIKHDAVS-----FRBLKWTSP-L 997
 752 REILDHSGKW----INKESEIIKDSSKEY-ISPFPKENKI-IVKSKYLHELSTLLQBR 805
 998 SNIVDDKSKWELWKNRLLKKEVEDEVEISEKIKETLNLNLESNYKEEIKYKLSFKY 1057
 806 NNANSSDI-----DLEKKVMTCEINAVANIDR--QIVEGR-----IEBAKNTLSDSI 852
 1058 KYDNASSIPTKLSVDVKQFIDKE-----NTBELFKLELRKPMFMEKKKISPSE 1112
 853 NIKNEP--KLTESIDSYDLKHQ-NGLDSDHFSIFEDISKTEGFR-RE----- 900
 1113 GTIHLFMQHLDLKKAENEDIEKQINRLIEREFITYEQ-SKVINPYKILKFCRGELGKR 1171
 901 -----INKETGSIETIEKETIFSEYATHISKE-----ISNKIDTIFDNVNGKLVKYNL 950
 1172 ILNNSNVKEMPPSIIPA-LEIYKELDKIYKDEKLIIOGVIDYCFEEDGLVLDYKT 1230
 951 DAAHEYNTLNSAFFIOSLIYNTKESLSNLSYAMKVQVYAGLFS 995
 1231 DYVNDIEEKVRIEQ--IKY--YEBALNRIITGKNVVKDYLYLFS 1271
 RESULT 15
 AAB18294
 ID AAB18294 standard; protein; 980 AA.
 AC AAB18294;
 XX 07-NOV-2000 (first entry)
 XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:152.
 DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX Plasmodium falciparum.
 OS Plasmodium falciparum.
 XX WO200025728-A2.
 XX 11-MAY-2000.
 XX 05-NOV-1999; 99WO-US026796.
 XX 05-NOV-1998; 98US-0107131P.
 XX (HOFF/) HOFFMAN S.
 XX (CARU/) CARUCCI D.
 XX (GARD/) GARDNER M.
 XX (VENT/) VENTER J C.
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX WPI; 2000-365347/31.
 XX
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection.
 XX Disclosure; Page 362-365; 577pp; English.
 PS
 XX The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
 CC (II) are useful for the development of vaccines against P. falciparum
 CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
 CC immunogens comprising the sequences of (I), are useful in the detection
 CC of infection with P. falciparum. Furthermore, (I) (especially when they
 CC are fusions or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins

CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification
 XX
 SQ Sequence 980 AA;
 Query Match 5.2%; Score 272; DB 3; Length 980;
 Best Local Similarity 20.1%; Pred. No. 2.4e-06;
 Matches 224; Conservative 182; Mismatches 391; Indels 320; Gaps 52;
 QY 6 KAOLQKVVY-VKPRIQDEYVAILNALAEYHNHSSVVEKYLKLDINNLTDNYLNTYK 64
 DB 39 KEKGEKVIYELKELDNDE--KIINDLKQNSY-----QVYKMDYKRENNLINEIN 89
 QY 65 KSGENKALKPKFMYLTMEV---LELKNSLTPV-EKMLHFIWIGQINDTAIVNQWD 120
 DB 90 K-----LKLFIENKMTVERGEMNKKLEEMKQKKNELINNLNIDISDELKNCIEQVNS 143
 QY 121 VNSDYT-----VKVFYDSNAFLINTLTKTI---VES-----ATNNTLES 156
 DB 144 VSRNMANVEKENIINELQILRMKN---DTMKRISKVFQEKVKLKFYLTINNDIFS 199
 QY 157 FRENLDPEFDYKFKRMEIIVDKOKHFD-----YVKSQIENPFPIDNIKI-T- 208
 DB 200 KNEKLD-----MOKLNDVNEKYKIVVECLNNYKTEHKEQIEKKIERI--NTLKQN 249
 QY 209 --YLSNEYSKDLALNKVIE-----ESLNKITANNNGDIRNEKFADED 250
 DB 250 YYYLKVEYDLKKNKELEKNIHKGKLEHLSHCYEENQKLEEEIKRRNSFIKNKDR--KID 307
 QY 251 LVRLYNQELVERWNLAAASDILRISMLKBDGGVLDVLDLPGIQDPLFKSINKPDSITNT 310
 DB 308 LITNIENELLKK-----KEINNIKMEKQVKNNEQLLKDIDENEKQNEHVYKLNQ- 360
 QY 311 SWEMIKLEAIMKYKEYIPGYTSKNPMDLDEEVQSFESALSSKSKSEIFLPLDDIKVSP 370
 DB 361 --ELIKRE-----LQNKISKIDIEFCCKEKEDK-----INKEDDL 394
 QY 371 LEVKIAPANNVINOALISLSDSYCSDLVINQIKNRYKILNLDNLNPSINEGTDFNTWKI 430
 DB 395 LEKKKCIEN---LKDELINIK-----KQMEDKQMTNE-MDLNLSKVVELNRLNKT 441
 QY 431 FSDKLASISNE-DNMFMKIKITNLYKGVFADPVDVSTINLSGPGVY--TCAYQ-----D 480
 DB 442 YEKNIVELNNELDVYKKNLDEBFLK---BEEKKNIDM---VYKIEVEIQIKERENE 494
 QY 481 LLMPKDNSTNTHLEPELRLNPEFPKTKISQLTQETISLWFSNQARAKSQPEEYKKGYPE 540
 DB 495 ISDLKKNQNLHLVAKNE-----ELNEXEIIKKNKYDK-EINMIIEQYNKKI-- 539
 QY 541 GALGEDDNLDPQNTVLDKDYVSKILSSMKTRNKYTHYI-----VOLQDKISYEASC 595
 DB 540 -----QEEKDMLNNKIKSMQDQTHKNQIIEEMQFENKELKRLKNVCDMMLOQLIKENEK 594
 QY 596 NLFPSK-DPVSSITLYOKNIEGSETAYVYVYVADAIEKE-IDKYRIPYQISNK----- 643
 DB 595 HMQEKVEEYKNNLKKQ-----DOELKNVIOEYDERIEIQNKEMEDIVNDC 640
 QY 644 -----RNILKTIPIGHKSEFNFTDTPANLDVDSLSSEIETILNLAQAKISPKYIENL 695
 DB 641 EKLQAKANNKKLT-----TATNANNNNMLMDENLKEKDKKINDLAKMDEKKEEINK 694
 QY 696 LG-----CNMPSYISAEETYP---GKLLKIKDRVSELNPSISQDSITVSA 739
 DB 695 LVEERSKLBHSHVKIQNEMSLIVEQNEKLKEEMGLSRIAID-MEEIKDMKEYEEXKK 753
 QY 740 NQVEVPIINEEGREILDHSGKWINKEESIINKDISKEVISFPNKENKIVKSKYLHELST 799

Db	754	NEERKQNEERKQNEERK--NEEKKQNEERK-----NEEKKLEKQHQFE----	804
Qy	800	LLQEIERNANSDDIDLEKKVMTCEINVASNIDROIVEGRIBEAKNLTSDSINYIKNEF	859
Db	805	-----EEKERM--EYEHQKEDR---KKDKKKKGHSDEEKYNKKE	842
Qy	860	KLIESISDSLYDLKHQNGLDDSHFISFEDISKTENGFRIRPINKETGNSIFITEKEIFS	919
Db	843	KTKKSSNILF-----DEEYIIQLEEL-----RDTGENCFI-----	873
Qy	920	EYATHISKEISNIKDTTFDNVNGKLVKKVNLDAHEVNTL-----	959
Db	874	-YLSLSKEL-----DVIINKLKSODALLN-DAFNKINLAITSWNIFNEENKEGDNITTV	927
Qy	960	-NSAFFIQSLIEYNTTKESLSNLSV--AMKVQVYQAQ	993
Db	928	ENTATEGNIITIDENTTEVENNEEVYKIFSVEKYDNL	964

Search completed: April 1, 2004, 16:41:35
Job time : 50.75 secs

Db 1599 NINVIDKYFTLVGKTNLGVVEFCIDNNKNIDYFGEMKTSKSTIFSGNGRNVVVEPI 1658
Qy 1662 YHDDSGNISSTVINSQKVLGIDRYVNVKVIAPNLYT 1700
Db 1659 YN-PDGEDISTDFSYEPYEGIDRYINKVLIAPDLYT 1696

RESULT 5
S55805
alpha-toxin - Clostridium novyi (ATCC 19402)
C:Species: Clostridium novyi
A:Variety: ATCC 19402
C:Date: 28-Oct-1996 #sequence revision 08-Nov-1996 #text change 05-Nov-1999
C:Accession: S55805; S71294; S71158; S44273; 140834; S44272
R:Hofmann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.
Mol. Gen. Genet. 247, 670-679, 1995
A:Title: Sequencing and analysis of the gene encoding the alpha-toxin of Clostridium novyi
A:Reference number: 140834; MUID:95342160; PMID:7616958
A:Accession: S55805
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2178 <HOF>
A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA88565.1; PID:g755724
A:Accession: S71294
A:Molecule type: Protein
A:Residues: 1-15 <HOW>
R:Hofmann, F.
submitted to the EMBL Data Library, March 1995
A:Reference number: S71158
A:Accession: S71158
A:Molecule type: DNA
A:Residues: 1-1179, 'LKV', 1183, 'LVTHIGE', 1191-2178 <HOS>
A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA88565.1; PID:g755724
R:Hofmann, F.; Habermann, E.; von Eichel-Streiber, C.
submitted to the EMBL Data Library, July 1993
A:Description: Sequence analysis of Clostridium novyi alpha-toxin: a member of the family
A:Reference number: S44272
A:Accession: S44273
A:Molecule type: DNA
A:Residues: 1-243, 1204-2178 <HOA>
A:Cross-references: EMBL:Z23281
C:Genetics: tcn-alpha
C:Superfamily: cpl repeat homology
C:Keywords: virulence factor
F:1880-1899/Domain: cpl repeat homology <CP12>

Query Match 24.4%; Score 2120.5; DB 2; Length 2178;
Best Local Similarity 31.3%; Pred. No. 1.1e-68;
Matches 552; Conservative 359; Mismatches 706; Indels 147; Gaps 46;

Qy 3 LVNKAQLOKVVYKFRIOEYVAIALNABEYH-NMSESSVVEKYLKXDINNLTNYLN 61
Db 2 LITREQLMKIASIFPKRKEPEYNILDALENFNRIEGTSVKETYSKLSKLNELVNYQT 61
Qy 62 TYKSGRNKALKPKFVLTMEVLEKNNSLTP--VEKNLHFIMTGGQINDTAINYNOMK 119
Db 62 KYPSSGNLALFNEDSLSYSELRLIKNSRTSTIASKNLSFIMTGGPISQSLSEYNNMK 121
Qy 120 DVNSDYTVKPYDSNAFLINTLKTIVESAATNNLTESFRNLNDPEPDYKFKRKRMBII 179
Db 122 MFNKDYNIRLPYDKNSLLVNTLKTATIQESSKVIIEQNQSLIDGTGHNKFKYSDRMKLI 181
Qy 180 YDKQKHFDYKSOIENPEFIIDNIILKYLSNYSKDLALNKIYESLNKITTANNND 239
Db 182 YRYKRELKMLYENMKQNN--VDDIIFNLSTNFKYDYGKLNQKNNKNNKMIAGATD 238
Qy 240 IRNLEKFADEDLVRLYNOLVERWNLAAASDILIRISMLKEDGGYVLDVDPGLQPDLPK 299
Db 239 I-NTENILTNLKSYYQELIQTNWLAASDILIRIALIKYGGYVCLDPLPGVNLSEFN 297
Qy 300 SINKPDSITNTSWEMIKLEAINKKEYIPGYTSKNFDMLEBEVQRFESALSSKSKSEI 359

Db 298 DISKPNQMSNYWEAAIFEALANEKLMNYPKYMEQVPSEKERILSFVRN-HINDL 356
Qy 360 FLPLDDIKVSPLEV-----KIAFANNVINOALISLKDSYCSDLVINQINQINRYKIANDNL 414
Db 357 ILPLGDIKISQLEILLRLKAATGKTKFNAFIISNDLSITLANLISQLENRYEILNSII 416
Qy 415 NP--SINEGTD--FNTMKIFSPDKLASISNEDNMFMWIKITNYLKVGFAPDPVSTNLSG 470
Db 417 QEFKICETVDSVINSVSELVLETTPKNLSDMDGSSFPQIIGYLSGFKPEVNSTVFFSG 476
Qy 471 PGVVTGAYQDLMFKDN-----STNHLLEPRLNPEFPKTKISQITEQETSLKSFNQ 524
Db 477 PNIYSSTATCTYFIKNTFDMLSQONQEIFEAS-NNLYFSK-----THDEFKSSMLLRS 529
Qy 525 ARAKSEFEYKGYFEGALGEDDNLDPANTVLDKDVSKKI--LSSMKTRNKEYIHVY 582
Db 530 NIASKEFKLKIYIGTLYNEDGLNPKWKTTSSELLKVIIEVNSTKIYENYDNLML 589
Qy 583 QLQGDKISYEASCNLFSPYSSILYQKNIIEG-SETAYVYVADAIEKEIDKRIPIVQIS 641
Db 590 QIQGDDISYESAVNVFGKNPKSIL---IQGVDDFANVFYFENGIVQSDNINNLSRFN 645
Qy 642 NKNVILKTFIIGHGKSEBNTDTFANLVDVLSSEI-----ETILNLAADISPKVIEINLL 696
Db 646 DIKKIKLTLIGHGENVFENPKLFGKTYNDLYTNIKPKLQHLEREGVILNKKYLKINIL 705
Qy 697 GCMNFSYSISAEFYPGKLLKIKDRVSELMP-SISODSITVGSANOVEVINEEGREIL 755
Db 706 GCMYFTPKVDINSTFVGKLFNKIS---RDLQPKGFSKNQLEISANKYATIRINREGKREVL 762
Qy 756 DHSGKWNKEBSIIKDISSEYISFNPKENKIIKSKYLHELSTLLOEIRNNANSSDIDL 815
Db 763 DYFGKVGSTDLIAEQISNKVYVYVNEVENTLSRVEQLNKAFAFKDI-----NSLIQT 817
Qy 816 EKKVMTCEINVASNTDRIQVEGRIBEAKNLTSDSI---NYIKNEFKLIESDSLSYD 871
Db 818 TNNQELKQSLVNTVADLTITLYSELLKEDIPFELDNQIKERIILNEISRLHDFSNIL 877
Qy 872 LKHQGLDDSHFISFEDISKTEGFRIRPFNKETGNSIFETEKEIIESEVATHISKISN 931
Db 878 FYQKNISNNMILFDSIIKEKYINVKLANKITGETSVIKTYSDSDSLNFTNKKYKIVDD 937
Qy 932 IKDITFDNNVNGKLVKKNLDAHEVNTINSAFFQSLIEYNTTKESLSNLSVAMKVQVYA 991
Db 938 IKGIIVKDIINGEFIKKADFEIQNPSSLNSAMQLLIDYKPYTEILTNMTVTSIKVQAYA 997
Qy 992 QLFSTGLNTITDASKVVELSTALDETIDLTPT-SEGLPIIATIIDGVSLGAALKELSET 1051
Db 998 QIFQISIGAIQEAETIVTIIIDALNANFNILSKLVGSSVASVIIDGINLIAALTELKQV 1057
Qy 1052 NDPLLRQIBIAKIGIMAVNLPAASTAVTSALG---IAGFSIILLVPLAGISAGIPSLVN 1108
Db 1058 KTNFERKLEIAKVGMYSIGFLESLSLGLGATAVSEILGIVISVPVAGILVGLPSLVN 1117
Qy 1109 NELLIQDKATKVIDYFKHISLAETEGATLDDKTIIMPQDDVLVLESDIDFNNSITLGCSE 1168
Db 1118 NTLVLGEYNGCILDYFSKFPYIVGNKPFES-IQDNIIPYDDIAITAITEILNFKTNKFKYGVAK 1176
Qy 1169 IWRAGSGSGHTLTDIDHFFSSPSITYRKPLWSTYDVNLNKKKEIDFSLKMLVLPNAPNR 1228
Db 1177 ISGFEGGGGHTYWGNIHDYFAPSLDHYIE-LSYFALKLNDTNLP-KGNVVLPSGLNK 1234
Qy 1229 VFGYEMGWTTPFRSLDNDGTLLDRIRDHY-----EGQFYWRYFAFIADALITLKPRIE 1283
Db 1235 VYKPEISAIAGANSQEGNGVEVNLIRNYVDSNGNTKFPWKYEAPP-EYSFSYMRVEYF 1293
Qy 1284 DYNVRINDGNTRSPFIVPITTEQIRKNLSYFSGGSGSYLSLSPYNNQINDLAVENDT 1343
Db 1294 DTKANVILDNENKTLIIPVLTIDENRNKISYEILGDDGQYVILPWNQTNINIVSNKDI 1353
Qy 1344 WVIDVNVKNTIIESDEIQKGLLENILSKLNEDNKIILNNHTINFINVGDINEENRFIS 1403

Db 1354 WNPDSYIVKESKIEDNKFVLDGFINNIFSTLKVSDNGFKIGKQFIS-----IKNTPRAIN 1409
 Qy 1404 LPSILEDINIIIEIDL-VSKYKLLSGNCKMLIENSSD-IOCKIDHIGFNGEHQKIP 1461
 Db 1410 LSGFINNNI-VIVSYIYNHEKNS-TIISDDLNDIKNPNLDNINYLIGLSISDNTIN 1469
 Qy 1462 YSVIDNETKNGFIDYSKEGLFTABFNSIESIRIYNPDNSNMLFYSGKOLKDIRIINK 1521
 Db 1469 CIVRNDVWEGKI-----FNEK-KLVFIQNELEHLVDS-----VNK 1506
 Qy 1522 GDVKLLIGN-----YFKDDMKVSLSTIEDT-----NTIKLGVYLDENGVAQI 1565
 Db 1507 -DSQYLNNPINNVKYGKGYVEGFTLNSNTENKYSLYIENKMKLGLYL-ESSVFKT 1564
 Qy 1566 LKPMNAKALNP-SNSLMNFLESINIK---NIFYNNLDPN---TBFILDTN--FIISGS 1616
 Db 1565 IQDKIYSKEKVDNDIILSLIKFTVNIQLCPFMIVSGVDNRRNYLEYMLSTNNKIING- 1623
 Qy 1617 NSIQGELICDDKNTQPVFINPKIKETSVLYVGNRQNLIVEPSVHLDDSGNISSTVIN 1676
 Db 1624 ---GYWE---ND-----FNNTKIVDF-----EKCNVIVSGGNKUNSEGLADT-ID 1662
 Qy 1677 PSOKLYLGIDRYVNVKVIAPNLYT 1700
 Db 1663 VLDKDLNL--YIDSVIIIPKVT 1684

RESULT 6
 T00296
 toxin B - Escherichia coli plasmid p0157
 C:Species: Escherichia coli
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 17-Nov-2000
 C:Accession: T00296; T42195
 R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota, S.; Shinagawa, H.
 DNA Ref. 5, 1-9, 1998
 A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic E. coli serotype O157:H7, substrain RIMD 050952
 A:Reference number: Z14127; MUID:98290540; PMID:9628576
 A:Accession: T00296
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3169 <NA>
 A:Cross-references: EMBL:AB011549; NID:94589740; PIDN:BAA31815.1; PID:93337056
 A:Experimental source: strain EHEC O157:H7, substrain RIMD 050952
 R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
 Nucleic Acids Res. 26, 4196-4204, 1998
 A:Title: The complete DNA sequence and analysis of the large virulence plasmid of Escherichia coli serotype O157:H7
 A:Reference number: Z22068; MUID:98391744; PMID:9722640
 A:Accession: T42195
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-236, 'D', 238-1887, 'I', 1889-3169 <BUR>
 A:Cross-references: EMBL:AF074613; PIDN:AAC70163.1
 A:Experimental source: strain EDL933; serotype O157:H7
 C:Genetics:
 A:Gene: toxB
 A:Genome: plasmid p0157
 A:Note: L7095
 C:Superfamily: Escherichia coli probable cytotoxin
 C:Keywords: cytotoxin

Query Match 5.9%; Score 507.5; DB 2; Length 3169;
 Best Local Similarity 20.08; Pred. No. 1.9e-10;
 Matches 420; Conservative 338; Mismatches 716; Indels 621; Gaps 105;

Qy 14 YVKFRI---QEDVEYVAINAL--EEVHNMGESSVFKYKLDKIDNNLTNYLTYKSGR 68
 Db 160 YIKRKTGAEDQTTTITQSILINELLNGVDRNTI-PFQKISELNDIIHSYENWQIKNSR 218
 Qy 69 -----NKALKKFKB-----YLTMEVLELKNNSLTPPEKNLHFTWIGG 105
 Db 219 KGIEILVKQELLSSLLNVNKNQKLSDNASKIINLLGIEYQSHKV-DIEPFIHAVVAG 277

Qy 106 QINDTAINYNQKDVNSDYTVKVFYDSNAF-----LINTLKT----- 144
 Db 278 APPDNTSYITAFINTYKDYTYLLWIDPNFAFGAAKFSGLKNIAMVYALMLRTRTNPHLA 337
 Qy 145 -----IVESATNTLE--SPREMLNDPEFY-----NKFYRKMELIYDQKHF 186
 Db 338 EEMNEVLTKIQNTIEFKETRELRKLEENYKSLTSETKEKFNVFLESIMIGMODNY 397
 Qy 187 IDYKSOIENPEFI-----IDNLIK--TVLSNEYSKDLEALNKYTESLUNKITANNGN- 238
 Db 398 FTYCISNGISNTDQISLDFLTNVKLSPVQDNDFKSTVEKNRDIIDLKNTISQKFGDR 457
 Qy 239 ----DIRNLKFADEDLVRLYNQELVERWNLAAASDILRISMLKEGQGVVLDYDILPGIQ 294
 Db 458 FQURDINTLSEFKKPDYFFYQZEMLLRWNYAASQVRINILKEYGIIYDTDILPAYS 517
 Qy 295 PDLFKSINKPDSINTNSWEMIKLEAIMKYK--EYIPG--YTSKNFOMLDEEVSORSFESAL 350
 Db 518 DKVSOIINE-KSDSKRPFEDLKLRRRISESILSIKGEKYSIKH-DGLDETTLNQLNNIL 575
 Qy 351 SSKSDKSEIFLPLDDIKVSPLEVKIAPANNSVI-----NQALISLKDSY 394
 Db 576 -SEIEK---LTIDDY--FKPVETKVRDTPKIFRYOKWTENTWNRGNNNFMFLTHRGSK 629
 Qy 395 CSDLVINOIKNRYKILANDNLNPSINEGTENTMTKIFSDKLASISNE-----DN 443
 Db 630 CIDFILSGOKQYLEL-QRIRDNISYNNLFYT-----EDLKSNNVAIGGIPAKKYLEH 683
 Qy 444 MFMKIKITNVLKVGFAPDVRSTINLSGP----- 471
 Db 684 GLF-----SEYRQDGTIPYVYSTLNISGPDIMRMQKKYKSLGRIGBHVHKDKNKLSDVNF 739
 Qy 472 -GVYTGAYQDLLMP---KDNSTNIHLLEPE-----LRNPEPKTKISQIT---EQEIT 517
 Db 740 LGVYASSKNKNSPNMLNPSVGINITPDDESSWAVRNNDINKILPEKINCHVPEKLP 799
 Qy 518 SLWFSNQARAKSQEEYKGVFEALGDDNLPQONTVLDKOVVSKILSSMKTRKEY 577
 Db 800 SLY-----YEIDRSRFFQGW---DNKSIKHVTEINKDLI--KDNILLTSSNID 843
 Qy 578 IHTIVOLQGDKI SYEASCNLFSPKDPYSILYQKNIEGSETAYVYVADAETKEIDKYRIP 637
 Db 844 VKLIUKI--DRELVAISKI-----DNPLALRSIRILOQLANYVTISNTEPENTINFI 895
 Qy 638 YQISNKNR-----IKLTFIGHGKSEFNTDTFANLVDVS-----LSSEIE 676
 Db 896 YDFYKQKQDLSAIAKL-----FSRNDADTKIIVWNSVMEKNVFLREVIS 941
 Qy 677 TILNAXADISPKYIEINLGCNMPYSISABETYPQKLLKIKDRVSELMPISQDSI- 735
 Db 942 CVLRKSKVD---SVINEN-----KKNLSKEDA--GALRDYAKLKMKELFSLDGGYK 989
 Qy 736 -TVSANQYEVRIINEEGKEIILDHSGKWINKEESIIDKISSKEYISFNP----- 782
 Db 990 KIITTNAY---IKERDKL-----SGIITYNIENSIISGHESEFDIIRSQHGWGDIJSTVEQF 1041
 Qy 783 KENKIIVKSKYLHELSTLLOIRNANGSSDIDLEKKVWLTCEINVAS-----NIDROIV 837
 Db 1042 KKEFEYVKSE--LSAKSIFDDIKNK-YITDPETKRVLYHQLDSDIKERIAFLDISHVAY 1099
 Qy 838 EGRIEAKNLATSDSINYIKNEFKLIESI-----SDSLYDLKHQNG 877
 Db 1100 PGSLEKLLQLS---GVVFSIDINIIAEVLLASGVSGHSHGVVYVAPSPDKULELRHRT 1155
 Qy 878 LDDSHFTSFEDISKTENGFRIRFINKETGNSI---FIETEKEISEVATHISKISINIKD 934
 Db 1156 KSNSEWI--EKITP---YVYDILSDNVSNVLRPPLSEEQKKILNDIKLEISKVS---E 1206
 Qy 935 TIFDWNVCKYKVNLDAAHEVNTINSAPFIQSILIEYNTTK-----ESLSNLSV--- 983
 Db 1207 QYFMKLTQKSSVIGIKYSVDFDRYENLFLUSLPINQMLTLPFMYRYFEMLYDIHIGIE 1266
 Qy 984 --AMKVQVYAQLFSTGLNTITDASKVELV-----STAL-DETID 1020

```

Db 1267 NKANREITYKFSNLNLDPLNDRVNLGLIKKYKYLSEIHRLTNTSTFADISIP 1326
Qy 1021 LLPTLSBGLPIATII-----DGVSLGAIAK-----ELSETNDPLL 1056
Db 1327 LLOTHI---CPSTIIKKTEYVGHQTNAMTVASWKPYDFNSLGAISIDKSVDPAL 1383
Qy 1057 ROEBE-AKIGMAVNLTAASATVTSAL-----GIASGFSIL 1092
Db 1384 HTIVEQAKYNLLSWNDFFYNTHASITWDTIARQHKSTNIEFHPQSLFFDRDSKGKGLGSL 1443
Qy 1093 LVPLAGISAGIPSLVNN-----ELLQDKATKVIDYFKHISLAETEGAFULLDDKIIMPQD 1148
Db 1444 YLDTGGYGGYQKLRHNDITASTIYQK-----YNDNLKLSNRDFFLAKTQRLTMSN 1497
Qy 1149 DVLSEIDFNNSITLKGCEIWAEGSGHTLTDI---DHFFSPSITYKXPLWSYDVL 1206
Db 1498 EL-----GNRLKNAQLEVLK---DPLTTEGILVQRRISLLITTEVHSLAQOIS 1547
Qy 1207 NIKK-----EKIDFSKOLMVLNAPNRVFGYEMGWTGPFSLNDNDGTLL 1251
Db 1548 SFWRVTDNFGCHDFSHLAQALTFIKNI-----TSNR-----NFSLYGSGI--- 1589
Qy 1252 DRIRHY-EGOFYWEYFAFIADALITKLPRYEDTNVRINLDG---NTRSFIVPITTEQ 1307
Db 1590 ---VKIYFSESNNWKYIKPLVQGSLLRDILYLTTPKISTSGGSLNIMGHLPV--- 1642
Qy 1308 IRKNSLYSF-YGSGYSLSLSPYNNMIDNLNVDNTWVIDVNVKNTIESDEIQ--- 1363
Db 1643 -----SFVIDIGV-----INGNRISBST---DVKNKIRSLKINGDILQHYI 1681
Qy 1364 -----KGELIENILSKLIENDKI-----ILN---NHTNFY 1392
Db 1682 NTHYLSERQTKIKDIVDFLGTQDNTIKVKLESIDKPISEIQPLHSILSRQKEHVRNLL 1741
Qy 1393 -GDINE-SNRF-----ISITFSILEDINI-----ITEI---DIVSKSYKILLSGNCMK 1435
Db 1742 SGLLDFSNKLRKQGLSLKTNVLSVNNPKESKINSIDTVEVTVDLQGLRYRVIDTRVIG 1801
Qy 1436 LI---ENSSDIOQKIDHIFNGEHOXKYPYVIDNETKNGFYDYSK---KEGFTAFBS 1489
Db 1802 LTFKEGINSLSFALBHM-----IDAIMSVIGLVQVARIKWDNIDISADHA 1848
Qy 1490 NE-SIIRNIYMPDSNNLIYSSKLDIRIINKGDVK-LLIGNYFKDDMKVLSFTIEDT 1547
Db 1849 GAVSDIKNIV---DKFLGGLITLTNNRVNPGVSGASLEGFTSSGLEVCAS----- 1897
Qy 1548 NTKLNGV---YLDENGVAQILKFNNNAKSAINTSNLSMNFLESINIKNIYFNMLD--- 1600
Db 1898 ---RMGGTAGRYL---SNVAKVIKL-----PLLDI---GINIMWSLYDSSLNHAKA 1938
Qy 1601 -FNIBFI---LDTNF-IISGNSISIGQFE---LICDKDKNIQPYFINKIETSVTLYV- 1650
Db 1939 TQCIYIYIAIDVFSSTINTALSIGAIAYPPLAIA---IVPITI-PSHEVKNYAVTVN 1992
Qy 1651 ---GNQNLIYVPSYHLD-----SGNISSTVINFSQKLYG---IDRYANKVII 1694
Db 1993 QINERHKUWLEAKYLDNGSAKVLNINATGIDLSNNQVLGNLYLDRENPPL 2047

```

RESULT 7
 T28676
 rhostry protein - Plasmodium yoelii (fragment)
 C:Species: Plasmodium yoelii
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
 C:Accession: T28676; A45521
 R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
 Mol. Biochem. Parasitol. 76, 329-332, 1996
 A:Title: Comparison of two members of a multigene family coding for high-molecular mass
 A:Reference number: Z20507; MUID:97077455; PMID:8920022
 A:Accession: T28676
 A:Status: preliminary;
 A:Molecule type: DNA

A:Residues: 1-2401 <SIN>
 A:Cross-references: EMBL:U36927; NID:G1041784; PID:G1041785; PIDN:AA441263.1
 R:Keep, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
 Mol. Biochem. Parasitol. 42, 241-246, 1990
 A:Title: Identification of the gene for a Plasmodium yoelii rhostry protein. Multiple c
 A:Reference number: A45521; MUID:91101660; PMID:2270106
 A:Accession: A45521
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 2260-2401 <KEE>
 A:Cross-references: GB:M34281

Query Match 5.4%; Score 468.5; DB 2; Length 2401;
 Best Local Similarity 20.2%; Pred. No. 3.3e-09;
 Matches 410; Conservative 333; Mismatches 686; Indels 596; Gaps 105;

```

Qy 6 KAOLQMYVYKFRIOEDYVAILNALEYHN--MSESSVVEKYKLDKINNLTNLTNTY 63
Db 118 KDTMKKIIILLIQYM-NEFKGLNDAMTKLNGEGISQKFVINNQIKQKFDKSTYDE- 171
Qy 64 KXSGENKALKKPK--EYLTMEVL-ELKNSLTPTVEKNLHFIWGGQINDTAINYNQKD 120
Db 172 KKEGFESSELEAKNWEKKLEIITELKKQNETVOLDIKIRELIKQIKD---IIEQKI 227
Qy 121 VN-----SDVTVKVYDVSNAFLINTLTKTIVESATNTNTLESFRENLDNDFDYNK 170
Db 228 VNDLKLKLNKKIKETIEKIEYKAV---DLKKEIKD--NVYID--ELAKEPPYQITK 279
Qy 171 FYRKMEIITYDKQHFIDYKSIQIEE--NPEFII---DNI-----IKTVLSNEY 214
Db 280 YIEKKNEIYNTIKSDPKIYVGDIEQLYNEMPSVVOESNIEHIEKNTEILTKTKIDNVY 339
Qy 215 S-----KDLKLNKYIEESLNTKANTANG-----NDIRNLEK----- 245
Db 340 NNITQNETETVKASHLKNIEETNNKLSSETIIDIYIYGEITNLTNLEDFKNKEKGLNKK 399
Qy 246 ---FADEDL-VKLYNQELVE---RWNLAASDILRISMLKE-DGGVYLD--VDILPGIOP 295
Db 400 IDEYAKENQVQNVKSNILEIKKHYN-----DQINIDNIKEKEAKQNDVQFKEHKMTIIP 454
Qy 296 DLPKSIKSDSITNTSWEMKLEAIMKYKEYIPGYTKNFMDLDBEVOFSFSSALSSKD 355
Db 455 NEMK-YQKP---STEIKMKDEFLSKNKY-----NDFDKYKKEVSEHNKFTELN 503
Qy 356 KSEIFPLDDIKVSPLEVKIAPANNVINQALISKDSY-----C--SDL 398
Db 504 K--IKTEVSEDEIKKYENKFN-DSKSLINETKKSIEEYQNINTLKKVDDYIKVCLNTNE 560
Qy 399 VINQIKRYKILNDNLNPSINEGTDPNTTKIYFSDKLASI----- 438
Db 561 LITNCHNKQTLKDLKNQIKTIKETSIDKIYTDKFENILTDKTELETKTGTGSLNHH 620
Qy 439 -----SNEDNNMFM-----IKITNYLKVGFPADVRSITNL 468
Db 621 ESNKKELLYFDLKANLGNKNEMLYKFNKEKAVEDIKKNV-----DINKIVN 673
Qy 469 SGPGVYTGAYQDILLMPKDNSTIHLLPELRNFPFPKTIQSOLTEQETLSLMSFNOARAK 528
Db 674 IETIYTSYN---INEDIENEIG-----KSELTANTKVLKVKANVTNL----- 715
Qy 529 SQFEYKKGYFEGALGEDDNLDPQNTVL--DKDYVSKILSSMKT-----RNKEYIHI 581
Db 716 NEIKELKDYDFQDFGKKEKNIKYPDENKINDIDITLNQIKDSIETLTETIKNSE--NHI 773
Qy 582 VQLQG--DKISYBASCNLSPKSDPYSSILYQKNIEGSETAYVYVADAETKEIDK--YRIP 637
Db 774 DEIKGQIDKLKVKYFNKTMFNEDP-----KEIEKKLENIV 807
Qy 638 YQISNKNIKLTFIGHGKSEFNTDTFANLDVDSLSSEIETILNLAKADISPKYIEINL-- 695
Db 808 EKIDKKKI-----YKIDKLNLEISKIEN-DKTSLE-KLKNINLSY 847
Qy 696 ---LGCNMF-----SYSISAEETYPGKL-LLKIKRVSELMPSISQD----- 733

```


Db 553 KIEKIHASINYSEETERIGINIESLNTKVEKV---KENVTLNKLKIKELKHVDFSD 609
Qy 504 -----PKTK-----ISQTEQBITSLWSPNQBARAKSQEYFKYFEGALGE 545
Db 610 FGKGNKYTKDKIKINDIMAVSQIDQHINGLDDI-QKSESYYSEMKE-----Q 660
Qy 546 DDNLDPAQNTVLDKDYYSKILSSMKTRNKEYTHYVQLGDKISYEASCNLSPKDPYSS 605
Db 661 INKLEKYSNTEISNDNVE-----GKKKQOI-IVTKIDKKNIYEEINKLUSE----- 707
Qy 606 ILYQKNTEGSETAYYYVADAIEKIDKYPYQISNKRNIKLTFIGHGKSEENTDTFAN 665
Db 708 ---ISKLEKNTS-----LEKVKDNLVSG-QNLGNLFLEQIDDEKKKAE-NTIKS 753
Qy 666 L-----DVSLSI-----SSRIETILNAKADISPKYIEINLLOCNMPSYSISABEYTPGKLL 717
Db 754 MEAVIDDLNLIKKSQBIETEMDI-KWDINKE-----MEALKISHDDO-----K 796
Qy 718 KIKRVSELMPSISQDSITVSAQYVRINEEGKREILDHSGWIN--KEESIIKDISK 775
Db 797 KCHDK-----SKN-----HKENISDIYDKSKIIQDPSRESINDIKK 835
Qy 776 EYISFNPKEN-----KIIIVKSK-YLHELSTLLQEIERNAN 809
Db 836 LQKNVSESONHNSINOCLNEVANINILKNIKKIILDKVKEYTSIEKKNKINDELN 895
Qy 810 SSDIDLEK---KVMLTEIEINVASNIDRQIVEGRIE-----AKNLTSDSINYIKNEFK-- 860
Db 896 NSEKVIKIEGLDLSKCRSKINSTLDDKDIDCEIKINVLKKNILNAETN-ITNHPKNA 954
Qy 861 -----LIESISDSYDLKHQGLDDSHFISFEDISK---TENGFRTFIN 902
Db 955 BEYNKIVLSNFFNEMADNKSQYILEIKKNGTND-HDYNIKELKSHKDKSNGYK-----T 1009
Qy 903 KETGNSIETEKELFESYATHI-----SKSISNLIKDTIFPNVAGKLVKKNLDAH 954
Db 1010 EADQKKAIQKNKELFEQYKEEVTVLNKKYAVELKKNKFKDT-KNDSKQIIEIK--DAH 1066
Qy 955 EWTNLSAFFIQL-----TEYNTTESLNSLVAMKVQVYAOQFSTGLNTITDAS 1005
Db 1067 NYCTLESKSEKKWNEIKNEKIHIEDEVANDKNKAIT-SIKVSEVPFKT---KIINKIN 1122
Qy 1006 KWVELVSTALDITDILLPTLSEGLPIATIIDGVSLGAAIKELSETNDPLLRQ----- 1058
Db 1123 EIRTSDDCLKETNDLEQIS-NLSIDTQETKLTEGKQLTKLELLESLEKQKKNIEDQ 1181
Qy 1059 -----EIEAKIGIMAVNLTAASTALVTSALGIASGFSILLVPLAGISAGIPSLVNNELL 1113
Db 1182 KXELDEVNSKI-----KNIENTYN----- 1200
Qy 1114 QDKATKVIDYFKHIS-LAETEGAPTLDDKLIMPDDILVLSIEDFNNSI-----TLG 1165
Db 1201 QHKKNYEIVGIVEKINEIAKTNKQIESTKELIKPTIOHIS--SFNANDLEGIDSDENLG 1258
Qy 1166 KCEIWEAEGGS-----GHTLTDIDHFFSPTSITRKPMLSIVYDLNLIKKEKIDFSKD 1218
Db 1259 K---YNTENGNIYEFIKSYNLITNYLETVSKESITYNQ-----IQNKRIDTQKE 1305
Qy 1219 LAMVLPNAPRVFGYEMGTPGFRSLDNDGTKLDIRDHYEGQFVWRYPAFIADALITKL 1278
Db 1306 L--LKNIEN-----VNKAQSYLDVIK--ENEF-----DRIVTHF 1335
Qy 1279 KPRYEDTNVRINDGNTRSFIVPVTTEQIRKNLSYSFVSGG-GSYSLSPYNNNIDLN 1337
Db 1336 KKKLNTVNDNF-----KN-EYSKVNEGFDNISINTVKNSTDEN 1374
Qy 1338 LVEN-----DTWVIDVDNVKNITTESDEIKGELIENILSKLNIENKNIILNHTINF 1391
Db 1375 SLNLINLOTKEMANIVNTYYSKYEAENIFRN--IPKLANLIK-----IKNSGIDL 1428
Qy 1392 YGDINESNRFISLTFISILEDINIIL-----EIDLVSYSKILLSGNCKMLIENSDDI 1443

Db 1429 SKDKIA--ILSYLDSKTEDTLFIPSPQKQKTYTYKISDSYSLILLD-----ILKKSQEL 1481
Qy 1444 QOKIDHIGFNGHQ--KIPIYSYIDNETKNGFYDYSKKEG-LPTAFESNESIIRNTYMP 1500
Db 1482 QKK-----EQTKLI---PENRRLYEKVQATNELRGTLSDLYKKEKILSEV--- 1526
Qy 1501 DSNLFIYSSKDLKDI-----RIKNG---DVKLLIGNYPKDDMKVLSLFTIEDTNT 1549
Db 1527 ---KLLLHKSNELNKLSNCFONYDTILESSEKQYDQVKEKSNNYKQEKELGIDFNVTDMEE 1583
Qy 1550 IKLNGVYLDENGVAQILKFMNNAKSALNTSNLSMNPLESINIKNIFYNNLDPNTEFILT 1609
Db 1584 -KFN-----NDIKVIBELENNYSSBENNILOSKOKLKELTNKFNAEIKKIDDKLIEK 1636
Qy 1610 NFITSGNSIGOFELICDKKNIOPY-----FINFKIKETSYTILYVGNRQNLIVEPSYHL 1664
Db 1637 NDLI-----DKLIETRKNCMLFTHTLAEFLKIKITDYSKFIESATKFSKEFLKYI 1687
Qy 1665 DDSGN-ISSTVINFSQKY-LYCIDRYV-NKVIIAPN 1697
Db 1688 GDTNSLNDIATLQLKYDLHQINKYVTSKLSDATN 1723
RESULT 9
F82885
hypothetical protein UU482 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82885
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mi
A:Reference number: A82870
A:Accession: F82885
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4688 <GLA>
A:Cross-references: GB:AE002145; GB:AF222894; NID:G6899476; PIDN:AAF30894.1; GSPDB:GN00
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: UU482
A:Genetic code: SGC3
Query Match 5.0%; Score 431.5; DB 2; Length 4688;
Best Local Similarity 19.1%; Pred. No. 1.7e-07;
Matches 420; Conservative 329; Mismatches 743; Indels 709; Gaps 103;
Qy 13 VYVKFRIODEVVAIINALE-EYHMSSESVVVEKYKLIKDNINLTDNYLNT---YKSGR 68
Db 403 LYEVIGIKKTNEVDVNLKQIPYHNS-----INNLSNALNTEPYQYTKNGD 449
Qy 69 NKALKKFKBYLTME-----VLELKNS-----LTPVEKNLHFIVGGQINDTAINYNQWKD 120
Db 450 INLIAKVAFYVYVQVYGFKDQNNQEHQILAKVKKDGTAFTGTALNNNSYSLDKIVS 509
Qy 121 VNSDYTVKVFYDSNAPLINTLKTIVESATNLTLSFR-----ENL----- 161
Db 510 VSNPQNLV---SNFDLTSKQKLIKPAANASVSDSTKTKQILENLNLDLINQKLVAFTV 566
Qy 162 --NDPFDYNNKFKRMEIYDKQ---KHFIDYKSCIEEN----- 197
Db 567 DNNDKEYVAVVANDQNNKIIFSDNLPKGYI-YHLAKVENNDLNKVINLNDPELKDKIID 625
Qy 198 -----PEFIIDN-----IKTYLSNEYSKDL--EALNKY-----IEESLN 230
Db 626 KRLNILLDSHPDFTYDNGNLEHTQLANDLDDLQKALNNANVKGIVVDQDGIHEID 685
Qy 231 KITANNNDIRNLEKFADEBLVR--LYN-QELVERNLAAASDILRISMLKEDGGYLDV 287
Db 686 VSDANGKVIPTKLANNDPTKPNITVTLKVVVLKQNNQPNIDLI SEEQLSGDNHISFK- 744
Qy 288 DILPGIOPDLFKSINKPD-----SITNWSMEMIKLEALMKYKEYIPGVTSKNFDMLEEVQ 343

Db 745 -----KPTTAKTKENDDYEIFSFPNPSLANKKIKLTKTIDD-----NNTNTKTVEAS 791
 Qy 344 RSFEALSCKSKSIFIP-----LDIKV-----SPL----- 371
 Db 792 IGLDGAIFKTDSDAIFAPDHRYTLTKLEADNKKVANIDEISPLDRIYVKNQKGNVANADN 851
 Qy 372 --EVKIAFANNSVI-----NOALISLSDSYCSDLVINQIKRY-----KI-- 409
 Db 852 KHEFKIPDKNKKLFAVYKDKNNNEIHPDKTDGKGVIVPNNNLFDPNKIYDFDKIVD 911
 Qy 410 LNDNLNPSINEGDTFNTMKIFSD-----KLASISN--EDMMFMKIKITNLYKVG 457
 Db 912 LNEYFNKTLIDRNSINKDVSALNDGVDQARKLVLPAPAVSNVTVNAINFQVNLFDNLKLS 971
 Qy 458 ----FAPDVRSTINISGPGVYTGAVQ-----DILMFKN 487
 Db 972 YNRQFALTIKAVNNDLTKYATYDPKTNKYKLNDFETHLDANTKYKVDVLELNNK 1031
 Qy 488 STNIHLLEPBLNFEPPKTKISQLEQETISLWSFNQARAKSQ-----FEYKGYPEG 541
 Db 1032 EKPIKLKDDVLNFETTSATTIN-----PPIWTKFDVVVKTNNDTTITFEIDDK---DN 1084
 Qy 542 ALGEDDNLDFQAVNTLVKDYKSKILSSMKTKNKEYIHYVQLQDKKISYEASCNLFSD 601
 Db 1085 ILKNDOKI-YAQUALMDDLLDQDVINPLVNFST-----NKIASINGLDLKGNS 1133
 Qy 602 PYSSILYQKNIEGSETAYYYVADAEIK-----EIDKTRIPYQISNKNENIKLTFIGHGK 655
 Db 1134 KYSI-----KNL-----YLNQDNQKVYLFKNDVTKYE-QHFTNPKINLSF---NK 1177
 Qy 656 SEBNTTTP---ANL-----DVDSLSSEIETILNLAADISPKY 690
 Db 1178 SAVEQDIFADHANLFIDYDQKLRIINEDVKIYYQNDINTKNEIQ--IGYKV-VANNK 1234
 Qy 691 IEINLKGCMFNSYSIAEETYPGKLLKIKDRVSELMPS-----ISDSITV 737
 Db 1235 IKPNLVG-----LAKETTYVKKLEALNKGASSIVNSEFOLLDPNTVFNSTSNKNTILV 1287
 Qy 738 SANQY-----EVRINEEGREILDHSGKI-NKEESIILKDISREY 777
 Db 1288 GLNSIDWANDQPIINAKINIGDDFQDNQOVKLIYVNDNKEIKSAVTLKQORNYQ 1347
 Qy 778 ISFNPKENKIIVSKY-----LHSLTLQEI-----RNNA-----NSSDIDLEKKVWL 821
 Db 1348 EFSNLKRLYTFSKIYVETNNTQTLHKLDLTHQPSINPNSNAVSLKNTNIEITKRIIV 1407
 Qy 822 TECEINVASNIDQIVEGRIE--EAKNL--TSDSINYIKNEPKLIESIDSLYDLKHQNG 877
 Db 1408 -----NNDQSLISAKIEVDIDNVLNTADKPNIVYQ-----LENSDANNLKLATN 1454
 Qy 878 L-----DDSHFISFEDI-----SKTENGFRIRFINKETGNSIFTETEKE 916
 Db 1455 AQVVVNNKKFLFDVLKINQYVIKEISFNSKPTNAY-FNFTNKNNTNIVSYDEQN 1513
 Qy 917 IFSEYATHISKEISINTKIDTFDNN-----GKLVKKNLDAHAHEVNTL----- 959
 Db 1514 -----KISLSNNINPTSKPQINKNDKSNVDVLDQVDKQLLANQYLR 1557
 Qy 960 -----NSAFTQSILIEYNTKES--LNSLVAMKVQVYALQFS-TGLNITIDASKVV 1008
 Db 1558 LKLQLNDNKTWTDPIFNNAKISPKLSNL-----IHNRAVELEGILYFDDQNSVN 1610
 Qy 1009 ELVSTALDEIDILPLISEGLPIATIIDGVSIGAAIKELS-----EFNDPLLRQ 1058
 Db 1611 DMTNNOISFNSKHKPKIEPEPSLTTI--NYDTNNAIKTVSAHNAQVHPFKLNTDRALEN 1668
 Qy 1059 E-----TEAKIGIMAVNLTAATAIVTSALGIASGFSIILVPLAGIS 1100
 Db 1669 DQIVEAVFAPTNLNDQKVVEAKLNNVTSFNSEGELEFNLSGRBETTYELIKVTFKNP 1728
 Qy 1101 AGIPSLVNN-----ELILQDKATKVIDYFKHISIAETEGAFTLDD----- 1141
 Db 1729 NKAYELLNKGVIPEYKNGSGAVEFTTQKFEHKHVIDVWSSTSTNTTQETKVIDGIQRA 1788

Qy 1142 ----KIIMPDDDLVISE-----IDFNNSNI-----TLGKCEI 1169
 Db 1789 WNNKKLELYVESNILGDPEIKTIVDNNNSVHLSPDKKEVNLVLANLKPGRRVSLKKINI 1848
 Qy 1170 WRAEGSGHTLDDI--DHFF-----SSPSITVRKPWLSYDVVNLKKEKIDFS 1216
 Db 1849 KEVDNGQDHEFVKBEINVWNSFDVNLQSEITASSVEEINDRAP-----DKLQNTTIKINLK 1903
 Qy 1217 KDLMLVLPNAPNRVFGVEMGWTFQFRSLDNDGDKLLDRI--RDHYEGOFYWRYPAFIADALI 1275
 Db 1904 DENDILK-----TNDIATYDNEQKVDALVKTNAQNKY-----LEALI 1943
 Qy 1276 TKLKPRVEDTNVRINDGNTRSPIVPITTEQIRKNLSYFSGSGSYSLSPYNNNI-- 1334
 Db 1944 TNL-----VFNKNVILKTEFKNLSQTF-----IKVGKNTNVI 1977
 Qy 1335 ---DLN--LVENDTWI-----DVP---NVVKN-----ITIESDEIQKGELENIISLKN 1376
 Db 1978 YDESNNLKLINNDFOIIGPLSTDVNSTQNVANNKHVISSTLDFKVNPHISKNLAFK 2037
 Qy 1377 IE-----DNKIILN--HTINPYGDINESNRFISLT-PSILEDINIIIEIDL 1420
 Db 2038 FQNINGEVVYSPILTNSSIIIVNNKNVINFILDNLAKSNQLYRLVDVYVYIDNN--DT 2092
 Qy 1421 VSKSYKILLSCMKLIE-----NSSDIQOKIDH--- 1449
 Db 2093 INDKNKVPKANNTVRIIDAPGKTTISKSNNTWNTSTSSQFEVINSDDGNEVLNLEA 2152
 Qy 1450 -IGNGEHOQYIPYSYIDNETKNGFIDYSKEGLETFABESNIESIIRNIYPPSNNLFY 1508
 Db 2153 TISFKKQOTLLTPKV--NIKKNN--KYLKQIINLEPENRYVLESILLAKPN--- 2203
 Qy 1509 SSKDLKDRIINKGDV--KLLIGNYFKDMKVSFTIEDTNTI--KLNQVYLDENGVAQ 1564
 Db 2204 KTKKPLVVEILNKDDISFOAGYKVIQIKSNQSPVDTKQRIKLDGIONAWNEKQL 2263
 Qy 1565 ILKPMNAKSAALNLSLMLNPLSINIKNF-YANLDPN-----IEFILDIN----- 1610
 Db 2264 EITYSAND---NSTKTAAIKLEKNKLEVEFELTNLEKNRYTFTKRIELINDNNTKTPFN 2319
 Qy 1611 -----FIISGNSIGQFEL--ICDKKN-IQPYFINFKIKETSYYLVGNRQNLIVE 1659
 Db 2320 KSDSIQDKFVLSNNQVGVNIIEIQDRVHNLNSAKIRFELNDLDNLV--SNDEQATIT 2377
 Qy 1660 PSYHLDSGNSISSTVINFOSKLYGI-----DRYVNVK 1692
 Db 2378 YNNQTTISAKV---ITQNKQKLEATSNLVNKNDTIINKI 2415

RESULT 10

G70163

hypotheical prctein BB0512 - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999

C:Accession: G70163

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whi-

son, D.; Peterson, J.; Karavague, A.R.; Quackenbush, S.; Salzberg, S.; Hanson, M.; Vugl-

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A>Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943; PMID:9403685

A:Accession: G70163

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2166 <KLE>

A:Cross-references: GB:A3001153; GB:AE000783; NID:G2688419; PIDN:AAC66876.1; PID:G2688,

A:Experimental source: strain B31

Query Match 4.9%; Score 422; DB 2; Length 2166;

Best Local Similarity 20.2%; Pred. No. 1.3e-07;

Matches 435; Conservative 333; Mismatches 677; Indels 704; Gaps 112;

RESULT 11

T18427
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18427
R;Lawson, D.; Bowman, S.; Bartell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18427
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-3724 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1
C;Genetics:
A;Introns: 307/1; 1545/2
A;Note: C0335c

Query Match 4.6%; Score 399; DB 2; Length 3724;
Best Local Similarity 19.7%; Pred. No. 1.9e-06;
Matches 402; Conservative 306; Mismatches 611; Indels 724; Gaps 109;
QY 20 QDEYYVAINA-----LEEYHNSSESVVEKYLKLDINNLDNYLNTYKKS-----G 67
DB 618 EDEYHVLKENLKEDANEYNDKEN---KNNKTEILK-SKYLENEKRTLEELKLRG 672
QY 68 RNKALKFKKELTWEVLEKXNSLTPVEKNLHPFIWGGQINDTAINYINQWKDVNSDYTV 127
DB 673 KNIIFKOEKY-----NSLGEV-----IINEIQINEENKINDIQOG--- 708
QY 128 KVFYDSNAFLNTLKKTVESA--TNTLESFRENLDPEFDYKNKYKEMEIYDKQKH 185
DB 709 -----NISKQKIIOSSRTNDTFNIKDISLD--DLEKEKRKK-----KSOH 748
QY 186 FIDYV---KSOEENPEFIIDNIKTY-----LSNEYSKDL-----EALNKYIEBSLKI 232
DB 749 FIDNLVKADKNEISENINKICDNNINNYDESINNYDESINNYDESINNYDESINNI 808
QY 233 TANGNDI--RNLEKFADELVLVYQ-----ELVERNLAAASDILRISMLKEDGG 282
DB 809 YDENINNYDENINNYDENINNYDEGINIKICDDNILENKNLKTINDIVQV-----EENN 864
QY 283 VYLDVDILPGIPLFKSINKPSITNTSWEMIKLEAMKYKEYIPGYTSKPFMDLDEEV 342
DB 865 ESIE-----KNELMISLND--INNTYMPK-----ENVDFINKI 898
QY 343 QRFSPESALSCKSKSEIFLPLDDIKVSPLEVKAFANNSVINOALISLKSQSDLVING 402
DB 899 KR--ESLL--KIDK-----NIK-----DNN--NDDEVIMDNFYENDFIIN- 932
QY 403 IKKRYKILNDLNP-SINEGTDFTNTMKIFSDKLAS---ISNEDNMMP-MIKITNYLKV 457
DB 933 --HMEITNKELDPLEINTQNEFTENLDIKKKYITNDHPNDADKMFYEMNKLIN----- 985
QY 458 FAPVVRSTINLSGGVGTAVQDILLMFKONSTNIHLEPELRNFPFPKTKISOLTEQET 517
DB 986 --KDMKN-----KEGEFTKDTTSGSLQSHKIKYKNGEEKHDKNNNEEKN 1030
QY 518 SLWFSNOARAKSQPEEYKGYFEGALGED-DNLDFQANTVLDKDYV-----SKIL-- 567
DB 1031 ILYDENQV-----YSVLSDHKIEQDIQDIHSIQTNICDNEINQINEENSKEGVRI 1082
QY 568 --SMKTKNKEYIHYIYVQLQD---KISYEASCNLFPSKDPYSSILYQKNTEGSETAYYY 622
DB 1083 SGTOMENKNMKNMKNMKNMKNMKNMKNMKNMKNMKNMKNMKNMKNMKNMKNMKNMKN 1127
QY 623 VADAEIK-EIDKYRIPQISNKNRIKLTFIGHGKSEFNTDTFANLDVLSSEIETILN- 680
DB 1128 --DMEKNDMEKN---DMENENNE-----NKSDIENENKNEYKNMKNMKNMKNMKN 1176
QY 681 -----LAKADISPKYIEINLLGCMFYSYSABETYPGKLL---LKIKDRVSELMFST 731
DB 1177 TIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENN 1231

QY 732 QDSITV---SANOQVEVRAINEEGREILDHSGKWINKEESI-KD-----ISKEY 777
DB 1232 KNNIKIEDDAKQINQVNNI-QNVKHYKQKNDLNNVD-IINDVLKSDDKFENMINSKE- 1289
QY 778 ISFNPKNKIIVSK-----YHLSLSTLQEIIRNANSSDIDLEKKVMLTCEINVASNID 833
DB 1290 --INIKDFKVDKNDICLNLNELNK--QSYNGNTTEKCIDIH-----DFDYVENKD 1337
QY 834 QVVEGRIEAEAKNLTSPSINYKNEFKLIESDSLYDLAKHQGLDSDSHFISPEDISKTE 893
DB 1338 ---VHDKIHEDRKEYCD-----ESLKYLPADDIKKMR-----SFINKKSKRE 1379
QY 894 NGPFIIRINKETG---NSIPIE-TEKEIFSEYATHISKE-ISNIKOTIPDNV-NGKLIV-K 946
DB 1380 NFLIISYNEEIPFLIKLKVLEIIEPSIILKNKHIEKQFLPLGKYINDNLKGEKCISS 1439
QY 947 KVNLDAAHE-VNTLNSA--FFIOSLIBYNTTK-----ESLSNLSVAMKVQVVAQLPSTG 997
DB 1440 KIILDLMEYNTIISKNGVGLVFNYYNIIQLHLDVFLHKLGLSTIKYINYVHFLNIN 1499
QY 998 LNTTIDASKVVELVSTALDETIDLLPTLSEGL-PIATII---DGVSLGAA----- 1044
DB 1500 LKAVND-QKQVNIYDILLDDQKIIVSNPFGMPFKFLIILKXHNIVSIHSGKHFNPKIY 1558
QY 1045 IKELSETNDPLL-----ROEIEAKIGIMAVNLTAATAIVTSALGIASGFSILLVPLAG 1098
DB 1559 RRLLSKFNDDKOIDSNNVKEVEEK-----DYPFKHISLAETGAFTL-- 1138
QY 1099 ISAGIPSVNNELIQLDKATKVI-----HSELPAWDKVEKYKNVKNQENDHKVCSDEYVLSIS--SEKYNVDG 1631
DB 1584 -----HSELPAWDKVEKYKNVKNQENDHKVCSDEYVLSIS--SEKYNVDG 1631
QY 1139 -----HSELPAWDKVEKYKNVKNQENDHKVCSDEYVLSIS--SEKYNVDG 1631
DB 1632 EDEQEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQ 1691
QY 1139 -----LDDKIIMPQDLVLS-EIDFNNSI-TLGKCEIWRAGGSGHT 1179
DB 1692 DDD 1751
QY 1180 -----LTDIDHFF--SPSITYRKPWLSIYDVLNKK-EKIDFSKOLMVL----- 1222
DB 1752 DESYVNLKNEINITYCTSPDKYK---NYN--NLKYCDDMEHEKDLMLKHLVS 1805
QY 1223 -----NAPNRVFGYEMGTFPGFRSLDNDGTLLDR-----IRDHYEGQPYWR 1265
DB 1806 LLKDEKENIKNNEGEE-----SIDN--IILDKMKSDQBYNENKDLNNEV-- 1852
QY 1266 YFAFIADALITKLKRYEDTVNRINLGNTRSFIVPVITTEQIRKLSYSFYGGSGSYSL 1325
DB 1853 -----IENLSEHLARHYDIEMIPIDIVGNIR--LYKLFIYNNMKNN--YSYLKNNPF-- 1901
QY 1326 SLSPYNNMIDNLVENDTWIDV-----DNVKNITIESDEIQKGLIENILSKLNIED 1379
DB 1902 ---PYKKGKRGAKGMLMINIKRHEYDNLMKK---KNGNEKTG---NNKKNKNNKED 1952
QY 1380 NKIIANHTINFYGDINESNEFISLTSILBEDIITIEDLVSKSYKIL---LSGNCWKL 1436
DB 1953 KNDYINNN 1991
QY 1437 IENSSDIOCKIDHIGFNGEHQ-----KYIPYSYIDNETKYNFYIDSKYKGLTAFBSN 1490
DB 1992 IKNNN-----FNLYHYWSLIFYKCPVSYIND---NKIEGKKK---FCVYKKN 2034
QY 1491 ESIIIRNIYMPSSNNLFI-----YSKOLKDI-----RIINKGDVKLL----- 1527
DB 2035 K-----YMLNNDHYMIIEFLKHPKPYAKKVKESVCSLVYFFNMDIQKRIIYLLTETE 2089
QY 1528 -----IGNYP-----KDDMKYS-----LSFTIEDTNTIKLN----- 1553
DB 2090 KLIIHIDIGKYQYFFLQSHQDKQHKNNKNNNNDDDDVDNKNIEBAHIFEIHRIEGKTIS 2149

Db 1282 RNLANEINNNKKKNDWCINENNNKIGDVCVKYEMSLTVKETSLSKYEQVKVDEKCS 1341
Qy 1481 -----EGLTAEFENE--SIIRNIYMPDSNNLFYSSKOLKIRIINKGVDKLLI 1528
Db 1342 QYDEIRFOYDEKCFQYDEINKKYGALLNI--NITKNWDSKVDNNEIISVDNKVEGI 1398
Qy 1529 GNYFK-----DDMKVLSFT--JEDTNTKLNQVYLDENGAQILKPM--- 1569
Db 1399 ANYLKOIPELNEEIRLKEINKISLLYSNELNEKNSYDINMKHQEQ-----LLFLEKT 1453
Qy 1570 ---NNAKSALNTNSLMN-----FLESINI-KNIFYNLLDPN---IEFILD 1608
Db 1454 NKNEEKIINLTSQSDAYKKSDESKLCAQFVDDVNIYGNISNNIRTNREYKYEEMFD 1513
Qy 1609 TNFIISGNSIGOFELICDKX--NIQPYEINFKIKET-----SYLYGNRONLIVE 1659
Db 1514 TNIEEKNMELSKYIHLLENPRCKMIYENENIKSSNKIIGLYNRYRYGLREDLCKE 1573
Qy 1660 ---PSYHLDDSGNISSTVINFSQKLYGIDRYNKKVII 1694
Db 1574 EIVPS-----KIGNISKNENNNKN-NTCDGVDEKVTI 1606

RESULT 13
D81702
adherence factor TC0439 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis McPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C;Accession: D81702
R;Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis McPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: D81702
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3225 <RET>
A;Cross-references: GB:AE002312; GB:AE002160; MID:g7190482; PIDN:AAF39293.1; PID:g719048
A;Experimental source: strain Nigg (McPn)
C;Genetics:
A;Gene: TC0439

Query Match 4.58; Score 391; DB 2; Length 3225;
Best Local Similarity 19.28; Pred. No. 3e-06;
Matches 374; Conservative 306; Mismatches 670; Indels 598; Gaps 97;

Qy 2 NLVNKAOLQKMYVYKFKIQDEYVAIL-NALEEVHMSE-----SSVVEKYLKLDINN 54
Db 196 NILEKUSLQ-----EQIKHNDNISLNKEAINRYSDLNRRKSKGOSLLVRQAKILD--E 249
Qy 55 LTDNLYNTYKSGRNKALKFKFVLTWEVLELKNNSLTPVEKNLHPITWIGGQINDTAINY 114
Db 250 ILSTKTEREASNSVMTTIKGFTHRV-----PYEKNIHGIWAGSPPEGTDEY 300
Qy 115 INQWKVNSDYTVKVFYDSNAP-----LINTLKTIVESANNVTELESFRENLDNPEFDYK 170
Db 301 IKLFLHTYPEFSFLFWDKTAYGAAKFSSTLKRIAFDAVN---SLREATPEPVKQFVQ 356
Qy 171 FYRGRMIIYDKQHFDYKSOIEENPEPIIDNIITYLSNYSKDL----- 218
Db 357 RYDK-LKKSVDTSRDFDE--KQLSEQLVELYDNY-----NRFSEIKQSNFVLLHEM 407
Qy 219 -----BALNKY-----JEESLNKI 232
Db 408 ITIODSPFNVCQKGVGAITDETIEVLEKVLKVEEDLSHYKETIKKNKESIEKLVEI 467
Qy 233 TANNG-----NDIRNLEKFADEDLRLYN--QBLVERWNLAASDILRISMLKDGGVY 284
Db 468 NDSTGRREVVKDRLDKSL--QDLTNSYVETEMLLRWNYAATDQLRYMLKEYGGIY 525
Qy 285 LDVDILPGIQDPLPKSN-----KPDSTINT-SWEMIKL-----BAIMKYKEYIPGYT 331

Db 526 TDLIDIMPOYSQDVLQIMDVGGSRFFEHDKRRTLTSLFAALKLGSQKQTTVSFEAKKAMT 585
Qy 332 SKNPFMDLDE-EVORSFESALSKSDKSEIPLPLDDIKV--SPLEVKI-----AFANNSVI 383
Db 586 LPFTLQDKSQISEIKYLETETQAKKSLFQPMQDMVTTVRDFMPILQRYHKWQTMVVRGL 645
Qy 384 NQALISLSDSYCSDIYINQIKRYKILNDLNAPSINEGTDENTMTKI FSDKLKASISNEDN 443
Db 646 NGLMMAKDSAVVDAVIARAAYDEM-RALRONVVSGEFFRSL-----GDLEHVREKN 699
Qy 444 M-MFMIKITNYL-----KVGFAFDVRSTINLSPGVYITGAYQDILLM--FKDNSTNI 491
Db 700 IGGVLAK--NYLGSULFFDFRQSDVPVIGALSTIGISP-----DIIMTMSDYFTNL 749
Qy 492 -----HLLPELERN-----FEPPKTKISQLT-----EQBIT-----SLWFSNQA 525
Db 750 GPVGEDFLYEGKLGKAFLGAYQAQKTPKGLTYDMLPLSIGANDVPADASTWCETRQ 809
Qy 526 RAKSQF-----EYKKGIFEGALGEDD-----NLDFAQ-----NTVLDK 559
Db 810 HCAAEILLSDSISSEDEHPKGIERRVNPNDFSKLWSKEAQGILLSSDFADLLPRFNLLIES 869
Qy 560 DYVSKILSSMKTANKEYIHVIQLOGDKISYEASCNLFPSKDPVSSILYQKNIEGSETAY 619
Db 870 SALDIHTLSAL---DRDIQHLFTKVQ-----KDPVASVA----- 900
Qy 620 YYYVADAEIEIDKVR-IPYQIGN-----KRNKILTFIIGHKSEF----- 658
Db 901 ---VPSLQQLAEIMIRAIPEIRNOVHILPEAQAHFEADWKALQLYLHSHQPTEVVIWY 957
Qy 659 ---NTDTFANLDVDSLSSEIETILNLAADISP-----KYIEINLLGC-----NMF 701
Db 958 SSTHTQIVFGKDLLAVERAAAKSL-MSDHPDSLITSYLKYKTQSHGLVTEFDQEDFF 1016
Qy 702 SYSIS-ABEYTPGKLLIKIDRVSELMPSISQDSITVSANQYEVIRINEEGREILDSH-G 759
Db 1017 ELMVDIAEPELHKQLLKIEQVN-----SGLYSHVEHSLG 1052
Qy 760 KW--INKEE-----SIIKOISSXEYISFNPKENKIIVKSKY-----LHEL 797
Db 1053 EWLKLSKEERKSKFLKILKETFOEBEEDSOQKHFWFEELYEKRHOVRKDPKAKTQEL 1112
Qy 798 STLLQETIRNNANSDDID-----LEKKVMLTECEINVASNIDRQIVEGRIEAKNLTSD 850
Db 1113 ITVTCB-SQRVOAQDIDITYPAHKPFYQDLMKDGVAFEDISVITKYL-----LASD 1161
Qy 851 SINVIKNEFKLIESISDLYD-LKHQGLDDSHFISFEDISKTEGNGPRIRF--INKETGN 907
Db 1162 GVSGIITTDPIFPFPPSKQLIDAMKQSLG-----EDFGEHLYTLQWVYDWLSKET-N 1211
Qy 908 SIFITEK-----EIFSEYATH-----ISKEISNIKDTIFDNVNGKLVKKNLDA 953
Db 1212 SVTSEQAQKQLFKLHEKLEGYTHDLLIPPIDGVSAL-GLRFSTEBEGKVSRLVTSIA 1270
Qy 954 HEVNTLNSAFIQS-----LIEYNTTKESLNSLVAMKVQVYAQLPSTGLNTITDASK 1006
Db 1271 PQVP--NSASVAMTSYLYGLFLITKDIOGRLTHEILKERLOTYGGAYFIN----- 1319
Qy 1007 VVELVSTALDETIDLLPTLSEGLPIIATIDGVSIGRAIKELSETNDPLLRQEI----- 1060
Db 1320 -----ESKIDVLLALSRRKAQI-SLIDAHKALTGFSFSEASLALLTGRMPGTSRV 1369
Qy 1061 ---EAKIGIMAVNLTAATAIVT---SALGTASGFSILLVPLAGISAGIPLSVNNELLIQ 1114
Db 1370 LSREVEFGPSAIVMEGATAIRAQSYDAVGLRKDFLL-----PHTVPSI--QSIIVEQ 1420
Qy 1115 DRATKVID-----YFGHI-----SLAETEGAFLLDDKIIMPODDIVLSEIDFNNSNITL 1166
Db 1421 AKYT-VLSWPEFYENHADKWNLANRFGA-----EDLSVHPQTFLYDTE-----GR 1465
Qy 1167 CEIMWRAEGSGG---HTLTDDIDHFFSFPSITYRKPLMSIYDVLNKKKIDFSDKDLMWL 1222

Db 1466 C-----MGLALLYLADD-----SVSYRL-----LQQLMFL 1492

Qy 1223 PNAPNVFGYEMGWTPGFRSLNDGKYLDRIDHVEGQYWFYAFIADALITKLPY 1282

Db 1493 ASLFDEQNRNIPUTPADQKFLNKGSLIEWLQFGNQQLQTEGFFHTLEWIPQLMKHF 1552

Qy 1283 EDTNVRINLQGNTRSPVITTEQIRKNLSYFGSGSYSLSPYNNIDNLVEND 1342

Db 1553 ASSTV-----KSLDITTPAHSLVLSLGMNFRVTDPNYGHDP--SLEAAITFLER- 1602

Qy 1343 TWIDVDNVV-----KNIT-----IESDEIQK-----GELTENLSKL 1375

Db 1603 --MYQVSPAVLERVGFDEKESVTSQKLVHSLSELSQNAVFASSDLGFTSRYPFTTLEEM 1660

Qy 1376 NIEDNKILANNHT--INFY-----GDINESNRFISLTFISILEDI--NI 1414

Db 1661 TVRGPITMNRQHTDWTATYKIGTVQGRIDSRSTRSDLA-----FLKINGDILEEFLTRT 1716

Qy 1415 IIEDLVSKSYKILLGNC--MKLIENSSDIQKIDHIGFNGEHQKYPYSVIDNETKY 1471

Db 1717 VLDSLVDELIOQLLTKHGLEFGTTLSFSPSIVETAIDHVS-----LQAVKTKSR 1767

Qy 1472 NGFTDYSKKEGLTAEFNSIESIIRNIYMPDSNNLFYSSKDLDIRINKGDKVLIGNY 1531

Db 1768 MHTILOSLGERIP-----KLPKNSGVQSDKT-----SIDRVQLVDESATSATIDFTV 1814

Qy 1532 PKD---DMKVSLSFTIED-----TNTIKLNGVYLDENGVAQILKPM-- 1569

Db 1815 IKDQRSQKSIITVGIESLAGSPKFSASMEHVIGTGVLDLD-LGMTVWSLVQVRLVEA 1873

Qy 1570 NNAKSALNTSLSMNP---LESINIKNI 1594

Db 1874 QGQKDALAVANLVNKLIALE-VSIGNV 1900

RESULT 14

Tl8501

hypothetical protein C0760c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: Tl8501

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A:Reference number: 218935

A:Accession: Tl8501

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3394 <LAW>

A:Cross-references: EMBL:Z98551; NID:el331903; PID:el331922; PIDN:CAB1140.1

C:Genetics:

A:Map position: 3

A:Note: C0760c

Query Match 4.4%; Score 382.5; DB 2; Length 3394;

Best Local Similarity 18.9%; Pred. No. 6.4e-06;

Matches 399; Conservative 319; Mismatches 701; Indels 687; Gaps 99;

Qy 6 KALQKQYVVKPRIOEDEYVAILNALESYH---NWSESSV-----VEKYLKLDINNLT-D- 57

Db 699 KYLEKLNELKNIYKD-----IEKKLDLHGVINKNKEDYIYLOVEKQTLKIVISSYDY 754

Qy 58 -----NYL-----NTYKSGRNKALKKFKELYLTWEVLKKNLSLTPVEKLNLFHWIGQI 107

Db 755 TKMSENIHIFQNTTWNKMLNVHSSKDYNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNI 814

Qy 108 NDTAINYNQMKDVNSD-----YTVK-----VF 130

Db 815 ENNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNI 874

Qy 131 YDSNAFLINTLKKTI-----VESATNTNTLESFRENLDPEFYKFRGMELIYDKQ 183

Db 875 KDEKKTSNIIKESISLKKKHTKSIINNNDN--NNEDDDDDNDMLSVMSNDDVINKR 932

Qy 184 KHFIDYYSQTEENPEFIIDNIITKTYLSNEYSKDLBALNKYIEESLNKITANNNGDIRNL 243

Db 933 K-----ENKLEEBHVSFSPSN-----NEYIAHSFNSILLQLSNYIFI 973

Qy 244 EKFADEDLRLYNOLVERWNLAAASILLISM-----LKEDGGVYLDVILPGIQDPLF 298

Db 974 E-----CKQMEYFNGSNLLSYVDYTTITIELFYRLKKYNNIFSIQIILGTQYPSIL 1024

Qy 299 KSINKPDSITNTSEMIMKLBALMKYKVIPIGYTSKDFMDLDEEVQSFESALSSKSD--- 355

Db 1025 QKLHGGIYCLDKNNKKQNDGDKSQDDDKKKKNDG--DNKSQEDDDGKKKNDGDN 1083

Qy 356 KSEIFPLDDIKVSPLEKIAFANNVINQALISLSDSYCSDLVINOIKRYKILNDNLN 415

Db 1084 KSQ-----EDD-----YGNKKKNDD--DDDDSYKIELIVDELNCKKQNTDEEL 1126

Qy 416 PSINEGTFNTMK---IFSOKLASISNEDNMFMKITN---YLKV-----GF 458

Db 1127 YELMKGSDFDKKYNFYLNHNFINNINIFSTIISFHIEDKYKYYERYFNLPMNF 1186

Qy 459 APDVRSTINLSPGV-----YTGAYQDLL-----MFKDNSTN-----JHLEPE 497

Db 1187 -----SNVLSFDLLIRRFDKILRLTKYEQLEENYEKIKKKNKEEYHLHACIKELMN 1241

Qy 498 LNPFPPTKISQTEQITSLWGFNQAR--AKSQFBEYK-KGYFEGALGEDDNLD--- 550

Db 1242 LERYNNEKIIDEQINEXEKINININEKYLILEKEYEYQNKNFINA--QIENLEREK 1299

Qy 551 FAQNTVLDRKYVSKKI-----LSSMKTNRKEY-HYIVQLQCDKISYBASCNLPSKDPYS 604

Db 1300 QLOEBELIIOKDMVKNLEKNCDDIKIYEKEQYLTLLQENKDSHNY-----LKDAFE 1352

Qy 605 SILYQKNIEGSETAYVYVADAIBEIDYRIPQISNKRNIKLTFIGHGSEFNTDFA 664

Db 1353 NLL--NL--NEKLYDH--DISLNKINTLWEEKENKKN---TF-----HMNLAVE 1396

Qy 665 NLDVDSLSEIETINLAKADISPKYIEINLLGCMFYSISAEETYPKGL----- 715

Db 1397 NNMLLKVKELQNKYNIIEKELNERIKOINFRNNVSLSDNRSTGSHQINNNMYN 1456

Qy 716 -----LLKIKDRVSEL-----MPSISQDSITVSANOYEVRINEE----- 749

Db 1457 NTHLGFMGASKINNINISLYSNMIMHSRGSIIKKNKEDAGNSTQARMNKNKSDTNIIN 1516

Qy 750 -----GKREILDHSGKWINKEE 766

Db 1517 NIHTNDINNNMNNINNTLINSINHLIYPPFPHNNVNSPKMGVMDVTLAG--VNRKD 1574

Qy 767 SIIDKISSKEYISPNKPKENKIIVSKYLHELSTLQELRNANSSDIDL--EKKYMLTEC 824

Db 1575 DFLNLEENEENSF--LEYEIRIKS-----LQELCDKSEILKIKGEKNILITCI 1623

Qy 825 EI-----NVAENIDR--QIVGRIEEAKVL-----TSDSINYI----- 855

Db 1624 ETWKCFCKKSEIEIRLKEICKQEKKBKFFLINKSNEDKLYINSLLCDEKDKYDIVV 1683

Qy 856 -----KNEF-KLIESISDSLYD---LKHQGLDDSHFISFEDISKTENGFRIRFINKE 904

Db 1684 KOIKNNWRNEIDKLANDINEKSYEIKLKHEN---NNLINEMNLIK-----NKE 1729

Qy 905 TGNISPIETEKIFSEYATHISKEISNIKDTIFDNNVNGKLVKKNLDAHAHVNTLNSAFP 964

Db 1730 TENNNTKQEE---DYIKLIKDKDTNIQNEYND-----LLEKYN-----EYVVKNNMLY 1775

Qy 965 IQSLIENYNTKESL-----SNLSVAMKVQVYA--QLFSTGLNTITDASKVVELYSTALDETI 1019

Db 1776 NDANVLLEKHEGEIFLLKENIKILOKNTYLNDFMNFQNNQNYDN-----NLLKRLDQLF 1830

Qy 1020 DLLPTLSEGLPIIATIIDGVSLGAALKELSETNDPLLRQ-----EIEAKIGIMAVNLTA 1073

Db 1831 NINQDLQKH-----DINQKHELSQKDYDIEIKERLAIKTIKINK 1870

Qy 1074 ASTAIVTSALGIASGFSILLVFLAGISAGIPLSVNNELIQDKATKVIDYFKHISLAETE 1133

Db 927 EKDER-----DKYLEASINDYMDDKKKKYDYSIESLRGSDKI---KNDQ 969
Qy 869 LYDLKHQGL-----DDSHFISFEDISKTEGFRIFINKETGNSIFETEKEIFSE 920
Db 970 IYQGHSSLLYYDNDNNDNNNNYSSSSNNHHYYILTNDKRLNWDNFNNLEINNS 1029
Qy 921 YATHISKE---ISNIKDT-----IPDNVNGKLVKKVKNLDAHEVNTLSAFTIQLIEY 971
Db 1030 QNKVIEKNLEYINNVLTKTSNYEQSNNTSKDEHNISDSKSKEDTLNLS--RKSSVEY 1087
Qy 972 NT-TKESLSNLSVAMKVQVVAQLPSTGLNTITDASKVVELVSTALDETIDLLTLSEGLP 1030
Db 1088 NKKILOSTSNKSLNGAYE--NNLFS-----GKKKKNGTVLKD-IEHINDIQKKYP 1135
Qy 1031 IIAITIDGVSLGAALKELSETNDP-LLRQBIKAGIGIMAVNLTAASTAIVTSALGIASGF 1089
Db 1136 -----EDLNINCNVKVIENEKHLPLBLE-----YNLVSSDEKF----- 1171
Qy 1090 SILLVPLAGISAGIPSLVNNELILODKATKVIDYFRHISLAETEGAFLLDDK---IIM 1145
Db 1172 -----GLNKIKDNNII-----YMKH-----QNVHENLYDDNQKKHILF 1204
Qy 1446 PQDDLVLSEIDFNNSITLGCETWRAEGSGCHTLTDDIDHFFSSPSITYRKPAWSIYDV 1205
Db 1205 DTNKNV--SIQRNNNINSVIKTHNYEVEKNNKQORNYD-----NFTCDKKKKIYYNI 1254
Qy 1206 LNIKKKIDFSKDLVLPNAPRVFGYEMGT-----PGRSLONDGPK-LLDRIRCHYEG 1260
Db 1255 IN--SDKDIYHNNIITYKNEKEGIGNHNLNNDKDI TNFELLKLDGVKEFLDTFKDSY-- 1310
Qy 1261 QYWRYPAFIADALITLKPREDTNVRINLDGNTRSFIVPVITTEQIRKNLSYSFYSG 1320
Db 1311 -----IDCHNKKENI----- 1320
Qy 1321 GSYLSLSPYNNITDLNVDNTWIDVDNVVKV-----ITIESDEIQGELIENILSKLN 1376
Db 1321 -----LNMVTKN-----KEDHQLIDVADKIFNETNMTMDNNKIYDDK-----N 1359
Qy 1377 IEDNKIILNNTINIFYGDI-----NESNPFISLTFSILEDINIIIEIDLVSYSKI-L 1428
Db 1360 VHEKK-CTHNDVIHNMIDILSTSINKNEENLF-----IDTYQKQNRIGD 1402
Qy 1429 LSGNCKLIENSDDIQKIDHIGFNGEHOKVIPYSYIDNETKYNG-----FIDYS 1478
Db 1403 IYMNPRINTQEDDDDDNNNNNNNNNNNNNNKILIFETYNQDQMLHNNKNNLEGTFFSDFI 1462
Qy 1479 KKEGLFTAEPFSGNES---IIRNIYMPDSNNLFYSSKOLKDIRIINKGDEVKLLIGNYFKDD 1535
Db 1463 EKKNKIKIKKNKESYHKIDESILSNEKN-----KVSLLIINNKKDS 1504
Qy 1536 MKVSLSFTEIDNTIKLVGVLDENGVAQILKFMNNAKSALNTSLSMNFLESINIKNIF 1595
Db 1505 SSVD-----NNKNN 1552
Qy 1596 YNNLPNIEFLDTNFI-----ISGNSISIQPELICDDKNIQPYFINEKIKETS YTL 1648
Db 1553 -NNNNNNDSFKONNLINNN 1611
Qy 1649 YVGNRQNL-----IVEP-----SVHLDSDGNSISSTWVNFOSKLYGIDR 1687
Db 1612 LINNHPNLNSMKRDTIEFLKINGKNKILKDLKIQOQVERKRYKIQMDQ 1664

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:44:46 ; Search time 56.875 Seconds
(without alignment)
7840.786 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700
Perfect score: 8675
Sequence: 1 MNLVKAQLQKVMYKFRIQ.....VLYGIDRVYKVIAPNLYT 1700

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1069545 seqs, 262320428 residues

Total number of hits satisfying chosen parameters: 1069545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6759.5	77.9	2366	14 US-10-011-366-10	Sequence 10, Appl
2	6759.5	77.9	2366	15 US-10-354-774-10	Sequence 10, Appl
3	6759.5	77.9	2366	15 US-10-271-012-10	Sequence 10, Appl
4	4467	51.5	2710	14 US-10-011-366-6	Sequence 6, Appl
5	4467	51.5	2710	15 US-10-354-774-6	Sequence 6, Appl
6	4467	51.5	2710	15 US-10-271-012-6	Sequence 6, Appl
7	2204	25.4	556	12 US-10-463-957-1	Sequence 1, Appl
8	2193	25.3	556	12 US-10-463-957-11	Sequence 11, Appl
9	2190	25.2	556	12 US-10-463-957-12	Sequence 12, Appl
10	2051.5	23.6	522	12 US-10-463-957-9	Sequence 9, Appl
11	1979	22.8	500	12 US-10-463-957-3	Sequence 3, Appl
12	1734	20.0	420	12 US-10-463-957-5	Sequence 5, Appl
13	769	8.9	170	12 US-10-463-957-7	Sequence 7, Appl
14	512.5	5.9	3169	14 US-10-114-170-257	Sequence 257, App
15	431.5	5.0	4688	12 US-10-282-122A-76865	Sequence 76865, A

16	347	4.0	6641	12 US-10-282-122A-70580	Sequence 70580, A
17	340.5	3.9	5005	12 US-10-282-122A-76871	Sequence 76871, A
18	336	3.9	4620	12 US-10-282-122A-68921	Sequence 68921, A
19	312	3.6	1639	14 US-10-087-464-10	Sequence 10, Appl
20	309.5	3.6	1957	15 US-10-369-493-2070	Sequence 2070, Ap
21	304	3.5	2184	14 US-10-304-095-6	Sequence 6, Appl
22	296	3.4	2017	15 US-10-452-024-146	Sequence 146, App
23	290	3.3	1182	12 US-10-282-122A-53445	Sequence 53445, A
24	289	3.3	2575	12 US-10-282-122A-72134	Sequence 72134, A
25	288	3.3	1875	15 US-10-369-493-22285	Sequence 22285, A
26	287	3.3	1610	14 US-10-155-533-9	Sequence 9, Appl
27	286.5	3.3	2046	12 US-10-282-122A-72132	Sequence 72132, A
28	285.5	3.3	1948	14 US-10-032-585-7611	Sequence 7611, Ap
29	281.5	3.2	1196	12 US-10-032-585-7646	Sequence 7646, Ap
30	281.5	3.2	1196	12 US-10-282-122A-52737	Sequence 52737, A
31	281	3.2	3692	12 US-10-282-122A-71235	Sequence 71235, A
32	279	3.2	1178	12 US-10-282-122A-52434	Sequence 52434, A
33	275.5	3.2	1279	12 US-10-282-122A-52455	Sequence 52455, A
34	274	3.2	1847	14 US-10-032-585-7538	Sequence 7538, Ap
35	273.5	3.2	1928	15 US-10-369-493-22025	Sequence 22025, A
36	271	3.1	1494	12 US-10-282-122A-47128	Sequence 47128, A
37	269.5	3.1	2167	9 US-09-801-368-56	Sequence 56, Appl
38	269.5	3.1	2167	15 US-10-369-493-1773	Sequence 1773, Ap
39	268.5	3.1	1184	12 US-10-282-122A-53254	Sequence 53254, A
40	267.5	3.1	1837	15 US-10-369-493-22734	Sequence 22734, A
41	267.5	3.1	6713	12 US-10-282-122A-43811	Sequence 43811, A
42	266.5	3.1	1465	12 US-10-282-122A-47271	Sequence 47271, A
43	264.5	3.0	3533	12 US-10-282-122A-70177	Sequence 70177, A
44	262	3.0	2013	12 US-10-282-122A-60608	Sequence 60608, A
45	259.5	3.0	1847	15 US-10-369-493-1075	Sequence 1075, Ap

ALIGNMENTS

RESULT 1

US-10-011-366-10
; Sequence 10, Application US/10011366
; Publication No. US20030054493A1
; GENERAL INFORMATION:

APPLICANT: Williams, James A.
Kink, John A.

TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES.

OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE

DISEASE

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/011,366

FILING DATE: 16-No. US20030054493A1-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/957,310

FILING DATE: 23-OCT-1997

APPLICATION NUMBER: US 08/329,154

FILING DATE: 24-OCT-1994

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-011-366-10

Query Match 77.9%; Score 6759.5; DB 14; Length 2366;
Best Local Similarity 76.6%; Pred. No. 0;
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;

QY	1	MNLVNAQLQKQVYKFRIOEDYVAILNALALEEYHNHNSSESVKYLKIDNNLTDNYL	60
DB	1	MSLVNRKLEKXANVRPTQDEYVAILDALALEEYHNHNSENTVVEKYLKIDNSLTDIYI	60
QY	61	NTYKSGRNKALKKPEYLTWVLELKNLSLTPVEKNLHPFIWIGQINDTAIYNOWKD	120
DB	61	DTYKSGRNKALKKPEYLTWVLELKNLSLTPVEKNLHPFIWIGQINDTAIYNOWKD	120
QY	121	VNSDYTVKVFYDSNAFLNTLTKTTVESATNTNLTSPRENLDPEFYKPKRMELIY	180
DB	121	VNSDYTVKVFYDSNAFLNTLTKTTVESATNTNLTSPRENLDPEFYKPKRMELIY	180
QY	181	DKQKHFIDYKQIBENPEFIDNIITKYLSEYKSKOLEALNKYIEBSLNKITANNNDI	240
DB	181	DKQKFINYKQARENPELIDDIIVKYLSEYKSKOLEALNKYIEBSLNKITANNNDI	240
QY	241	RNLEKPAEDLVRLVQBELVERNLAAASDIIRLSMLKDGCVGLVDVILPGIQDPLFKS	300
DB	241	RNLEKPAEDLVRLVQBELVERNLAAASDIIRLSMLKDGCVGLVDVILPGIQDPLFKS	300
QY	301	INKPSITNTSHMILKLAIMKYKVIIPGYTSKQFMDLDEEVQSFESALSKDKSEIF	360
DB	301	INKPSITNTSHMILKLAIMKYKVIIPGYTSKQFMDLDEEVQSFESALSKDKSEIF	360
QY	361	LPDLIDKVPLEKVLAFANNSVINQALSLKDSYCSLDLVINQIKRYKILNDNLNPSINE	420
DB	361	LPDLIDKVPLEKVLAFANNSVINQALSLKDSYCSLDLVINQIKRYKILNDNLNPSINE	420
QY	421	GTDFNTMTKIFSDKLASISNEDNMFMKITNYLVKGFAPDVRSTINLSGPGVYTGAYQD	480
DB	421	GTDFNTMTKIFSDKLASISNEDNMFMKITNYLVKGFAPDVRSTINLSGPGVYTGAYQD	480
QY	481	LLMFKDNSTNHLBPENRFPKTKISQLTQEBITSLWFPNQARAKSQPEEYKGYFE	540
DB	481	LLMFKDNSTNHLBPENRFPKTKISQLTQEBITSLWFPNQARAKSQPEEYKGYFE	540
QY	541	GALGEDDNLDAQNTVLDKDYVSKILSMKTRKEYTHYIYVLOGDKISYEASCNLFPSK	600
DB	541	GALGEDDNLDAQNTVLDKDYVSKILSMKTRKEYTHYIYVLOGDKISYEASCNLFPSK	600
QY	601	DPYSSILYQKNEGSETAYYVVADEAIKEIDKIRIPYQISNKRNIKLTIFGHGKSEPT	660
DB	601	DPYSSILYQKNEGSETAYYVVADEAIKEIDKIRIPYQISNKRNIKLTIFGHGKSEPT	660
QY	661	DTFANLVDVLSSEITETINLAKADISPKYIEINLLGNMFPYSISARETYPGKLLLIK	720
DB	661	DTFANLVDVLSSEITETINLAKADISPKYIEINLLGNMFPYSISARETYPGKLLLIK	720
QY	721	DRVSELMPSISQDSITVSANQEVRIINEEKEILHDHSGKWNKEEIIKDISKEYISF	780
DB	721	DRVSELMPSISQDSITVSANQEVRIINEEKEILHDHSGKWNKEEIIKDISKEYISF	780
QY	781	NPKENKIIVKSRYLHELSTLLQEIARNNSSDIDLEKKVMTCEINVASNIDRQIVBGR	840

DB	781	NPKENKIIVKSRYLHELSTLLQEIARNNSSDIDLEKKVMTCEINVASNIDRQIVBGR	840
QY	841	IEEAKNLTSDSINIVKNEFKLIESISDLYDLKHONGLDSDSHFISFEDISKTENGFRIRF	900
DB	841	IEEAKNLTSDSINIVKNEFKLIESISDLYDLKHONGLDSDSHFISFEDISKTENGFRIRF	900
QY	901	INKETGNSIFITEKEIPSEYATHISKEISNIKDTIFDNVNGKLVKKVNDAAHEVNTLN	960
DB	901	INKETGNSIFITEKEIPSEYATHISKEISNIKDTIFDNVNGKLVKKVNDAAHEVNTLN	960
QY	961	SAFFIQSILIEYNTTKYESLSNVAMKQVYAQLSTGLNTITDASKVVELSTALDETID	1020
DB	961	SAFFIQSILIEYNTTKYESLSNVAMKQVYAQLSTGLNTITDASKVVELSTALDETID	1020
QY	1021	LLPTLSEGLPIIATIIDGVSIGAAIKELSETNDPLLRQIEIAKIGIYAVNLTAATASTVT	1080
DB	1021	LLPTLSEGLPIIATIIDGVSIGAAIKELSETNDPLLRQIEIAKIGIYAVNLTAATASTVT	1080
QY	1081	SALGTASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGAFTLLD	1140
DB	1081	SALGTASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGAFTLLD	1140
QY	1141	DKIIMPQDDLVSEIDFNNNSITLCKEIRAEAGSGHTLTDIDHFFSSSITYRKFWL	1200
DB	1141	DKIIMPQDDLVSEIDFNNNSITLCKEIRAEAGSGHTLTDIDHFFSSSITYRKFWL	1200
QY	1201	SIYDVNLTKKEIKIDFSKOLMVLPNAPNRFVGVEMGTGPFSLDNDGKLDRIEDHYEG	1260
DB	1201	SIYDVNLTKKEIKIDFSKOLMVLPNAPNRFVGVEMGTGPFSLDNDGKLDRIEDHYEG	1260
QY	1261	QFVRYFAFIADALLTKLPYEDTNVRLINDGNTRSFIVPVTTEQIRKNSLYSFYSG	1320
DB	1261	QFVRYFAFIADALLTKLPYEDTNVRLINDGNTRSFIVPVTTEQIRKNSLYSFYSG	1320
QY	1321	GSYLSLSLSPYNNKIDNLVENDTWIVDNNVKNITIESDEIQKGLIENILSKLNIEDN	1380
DB	1321	GSYLSLSLSPYNNKIDNLVENDTWIVDNNVKNITIESDEIQKGLIENILSKLNIEDN	1380
QY	1381	KILNNHTNFYGDINESNRFSLTFSILEIDNIIIEIDLVSYSKILLSGCMKLIENS	1440
DB	1381	KILNNHTNFYGDINESNRFSLTFSILEIDNIIIEIDLVSYSKILLSGCMKLIENS	1440
QY	1441	SDIQOKIDHIGNGEHOKYIPYSYIDNETKYNFIDYSKKEGLFTAEFNSESIIRNIYMP	1500
DB	1441	SDIQOKIDHIGNGEHOKYIPYSYIDNETKYNFIDYSKKEGLFTAEFNSESIIRNIYMP	1500
QY	1501	DSNNLFYSSKDKIDRIINKGDVKLLIGNYFKDMKVSLSFTIEDTNTIKLNGVYLDEN	1560
DB	1501	DSNNLFYSSKDKIDRIINKGDVKLLIGNYFKDMKVSLSFTIEDTNTIKLNGVYLDEN	1560
QY	1561	GVAQILKFMNNAKSALNTSNLMNLESINIKNIFVNNLDPNIEFILDNFIISGNSIG	1620
DB	1561	GVAQILKFMNNAKSALNTSNLMNLESINIKNIFVNNLDPNIEFILDNFIISGNSIG	1620
QY	1621	QFELICDKKNQIOPYINFKIKETSTLYVGNRQNLIVEPSYHLDSDGNSISSTVNFOSK	1680
DB	1621	QFELICDKKNQIOPYINFKIKETSTLYVGNRQNLIVEPSYHLDSDGNSISSTVNFOSK	1680
QY	1681	XYGIDRYNKKVIAANLAYT 1700	
DB	1681	XYGIDRYNKKVIAANLAYT 1700	
QY	1700	XYGIDSCVKNKVVISPNIT 1699	
DB	1700	XYGIDSCVKNKVVISPNIT 1699	

RESULT 2
US-10-354-774-10
Sequence 10, Application US/10354774
Publication No. US20030215468A1
GENERAL INFORMATION:
APPLICANT: Williams, James A.
; Thalley, Bruce S.
TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin

NUMBER OF SEQUENCES: 82
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/354,774
FILING DATE: 30-Jan-2003
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/704,159
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-02304
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-354-774-10

Query Match 77.9%; Score 6759.5; DB 15; Length 2366;

Best Local Similarity 76.6%; Pred. No. 0;

Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;

QY 1 MNLVKAQLOQWVYKFRIOSEVVAIINALAEVHNHNSSESVVEKYLKLDINLTNDYL 60
DB 1 MSLVNRKQLEKAAVRPTQDEVAIDALAEVHNHNSSESVVEKYLKLDINLTNDYL 60
QY 61 NTYKSGRNKALKKFKELYTMEVLELKNLSLTPVEKNLHPITWIGGQINDTAINYNQW 120
DB 61 DTYKSGRNKALKKFKELYTMEVLELKNLSLTPVEKNLHPITWIGGQINDTAINYNQW 120
QY 121 VNSDYTVKVDNAPLINTLKTIVESATNTLESFRENLDNPEFDYKFRKRMEIY 180
DB 121 VNSDYTVKVDNAPLINTLKTIVESATNTLESFRENLDNPEFDYKFRKRMEIY 180
QY 181 DKQKHFIDYKSOJTEENPFTIDNIKTLYLSNEYSKDLEALNKYIEESLNKITTANNNDI 240
DB 181 DKQKFNINYKAQREAPELIIDIVKTYLSNEYSKEIDELNTYIEESLNKITTANNNDI 240
QY 241 RNLEKFADEDLRLYNQELVERMNLAAASDILRISMLKEDGGVYLDVILPGIQDPLFS 300
DB 241 RNLEKFADEDLRLYNQELVERMNLAAASDILRISMLKEDGGVYLDVILPGIQDPLFS 300
QY 301 INKPDSTNTWEMIKLEALNKYIEYIPGYTSKNFMDLEVOVSFESALSKSKSEIF 360
DB 301 IEKPSVTVDFWEMTKLEALNKYIEYIPGYTSKNFMDLEVOVSFESALSKSKSEIF 360
QY 361 LPLDDIKVPLEVIAFANNVINOALISLKDSYCSPLVINQIKNRYKIINDNLNPSINE 420
DB 361 SSGDMEASPLEVIAFANNVINOALISLKDSYCSPLVINQIKNRYKIINDNLNPSINE 420
QY 421 GTDFNTTKIFSDKLASINEDNMFMKIKNYLVKGFAPDVRSTINLSPGYITGAYOD 480
DB 421 DNDFTNTNTFIDSIMAEANADNGRFMMELGKYLVRGVFFDVKTINLSGPEAYAAAYOD 480
QY 481 LLMFKDNGSTNTHLLEPELRNPEFPKTKISQLTQEITSLMSFNQARAKQFEEYKGYFE 540

DB 481 LLMFKEGSMNHLIEADLRFNFEISKNTISQTEQEMASLWSFDDARAKAQFEEYKRYNYPE 540
QY 541 GAIGEDDNLDPACNTVLKDYVSKYLSSMKTKENKEYIHYIVLOQGDKISYEASCNLPSK 600
DB 541 GSLGEDDNLDPACNTVLKDYVSKYLSSMKTKENKEYIHYIVLOQGDKISYEASCNLPSK 600
QY 601 DPXSSILYQKNIEGSETAYYYYVADAEIKEDIKYRIPYQISNKRNIKLTFFIGHKSEFNT 660
DB 601 TPYDSVLFOKNIEDSEIAYVYVPGDGEIQEIDYKIPSIISDRPKIKLTFIGHKXDFNT 660
QY 661 DTPANLVDLSSEIEITILNLAADISPKYIEINLLGCMNFSYSISAEITYPKLLKLIK 720
DB 661 DIFAGFDVDSLSTEIEAADIADLAKEDISPKSIEINLLGCMNFSYSINVEITYPKLLKLIK 720
QY 721 DRVSELMPSISODSITVSANQYEVVRNEEGKEIILDSGWINKKEESIIDKSSKEYISF 780
DB 721 DKISELMPSISODSITVSANQYEVVRINSEGRRELLDHSGEWINKKEESIIDKSSKEYISF 780
QY 781 NPKENKIIVKSKYLHELSTLLQEIARNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840
DB 781 NPKENKITVSKNLPESLSTLLQEIARNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840
QY 841 ISEAKNLTSDSINYIKNEPKLIESISDSLYDLKHONGLDSDSHFISPEDISKTENGEPF 900
DB 841 ISEAKNLTSDSINYIKNEPKLIESISDSLYDLKHONGLDSDSHFISPEDISKTENGEPF 900
QY 901 INKETGSIETETEKEIFSEYATHISKEISNTKDTIFDNVNGKLVKKNLDAAEVNTLN 960
DB 901 INKETGSIETETEKEIFSEYATHISKEISNTKDTIFDNVNGKLVKKNLDAAEVNTLN 960
QY 961 SAPFIQSLIEYNTKESLSNLSVAMQVYQAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAPFIQSLIEYNTKESLSNLSVAMQVYQAQLFSTGLNTITDASKVVELVSTALDETID 1020
QY 1021 LLPTLSGLPIIATIIDGVSGLGNAIKELSETNDPLLRQIEAKIGIMAVNLTAATAIYT 1080
DB 1021 LLPTLSGLPIIATIIDGVSGLGNAIKELSETNDPLLRQIEAKIGIMAVNLTAATAIYT 1080
QY 1081 SALGIASGPIILLVPLAGISAGIPSVNNELILODKATKVIDYFKHISLAETEGAFITLD 1140
DB 1081 SSGIASGPIILLVPLAGISAGIPSVNNELVLRDARKATKVYFVKHVSLEVEGFTILD 1140
QY 1141 DKIMPODDIVLSEIDFNNNNSITLGRKEIWRAGGGSGHTLTDIDHFFSPSITYRKPLW 1200
DB 1141 DKIMPODDIVLSEIDFNNNNSITLGRKEIWRAGGGSGHTLTDIDHFFSPSITYRKPLW 1200
QY 1201 SIYDVNLKEKIDFSKDLMLPNAPRVFGYEMGWTPGFRSLDNDGTLLDRIDRHEG 1260
DB 1201 SIYDVNLKEKIDFSKDLMLPNAPRVFGYEMGWTPGFRSLDNDGTLLDRIDRHEG 1260
QY 1261 QFYWRYPFAFIADALITKLPYVEDTNRINLGNTPSFIVPTTBOIRKNSLSYFYGSG 1320
DB 1261 QFYWRYPFAFIADALITKLPYVEDTNRINLGNTPSFIVPTTBOIRKNSLSYFYGSG 1320
QY 1321 GSYLSLSPYNNMIDNLVENDTWIVDVNVKNITTESDEIQKGBELIENILSKLINEN 1380
DB 1321 GTVALSLSQNVGINTSELSDEWIIDVNVVRDVTIESDKIKGDLIEGILSTLSIEN 1380
QY 1381 KIILNHTINPYGDINESRPFISLTSLIEDINIIIEIDIVSKSYKILLSGNCMLIENS 1440
DB 1381 KIILNHTINPYGDINESRPFISLTSLIEDINIIIEIDIVSKSYKILLSGNCMLIENS 1440
QY 1441 SDIOQKIDHIGFNGEKGKIPYSDIDNETKNGFIDYSKKEGLTAAEFENESIIRNIYMP 1500
DB 1441 NHIOQKIDHIGFNGEKGKIPYSDIDNETKNGFIDYSKKEGLTAAEFENESIIRNIYMP 1500
QY 1501 DSNMLFTYSSKDLKDIRIINKGDVLLIGNYFKODMKVSLSTIEDTNTIKNGVYLDEN 1560
DB 1501 DSKPSFGYSSNLLKDVKITKDNVNLITGVYKDDIKISLSLTLODEKTIKLSVHLDSE 1560
QY 1561 GVAQILKFMNNAKSLNTSNLNMNLESINIKNIFANNLDPNIEPTLIDNFIISGNSIG 1620

Db	1561	GVABILKPMNR-KGNTWTSLSMFLSMNIXSIFVNFQCSNFKFILDANFIISGTTSG	1619
Qy	1621	QFELICDKDNIOPIYFINKETSVLYVGNQNIIVFSPYHLDDSGNISSTVINFSOK	1680
Db	1620	QFEFICDENIOPIYFKNTLTNTVLYVGNQNIIVFSPYHLDDSGNISSTVINFSOK	1679
Qy	1681	YLYGIDRYVKNVLIAPNLYT	1700
Db	1680	YLYGIDSCVKNWISPNYT	1699
RESULT 3			
US-10-271-012-10			
; Sequence 10, Application US/10271012			
; Publication No. US20030219457A1			
GENERAL INFORMATION:			
; APPLICANT: Williams, James A.			
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium			
; Botulinum Neurotoxin			
; NUMBER OF SEQUENCES: 82			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Medlen & Carroll			
; STREET: 220 Montgomery Street, Suite 2200			
; City: San Francisco			
; COUNTRY: California			
; COUNTRY: United States of America			
; ZIP: 94104			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Patent In Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/10/271,012			
; FILING DATE: 15-Oct-2002			
; CLASSIFICATION: <Unknown>			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US/08/704,159			
; FILING DATE: <Unknown>			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Ingolia, Diane E.			
; REGISTRATION NUMBER: 40,027			
; REFERENCE/DOCKET NUMBER: OPHD-02304			
TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (415) 705-8410			
; TELEFAX: (415) 397-8338			
; INFORMATION FOR SEQ ID NO: 10:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 2366 amino acids			
; TYPE: amino acid			
; TOPOLOGY: linear			
; MOLECULE TYPE: protein			
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:			
US-10-271-012-10			
Query Match 77.9%; Score 6759.5; DB 15; Length 2366;			
Best Local Similarity 76.6%; Pred. No. 0;			
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;			
Qy	1	MSLVNKAOLQKVVYKFRIOEDYVAILNALDEYHNMSESSVVEKYLKLDINNLTDNYL	60
Db	1	MSLVNKRQLEKMANVFRIOEDYVAILDALDEYHNMSENTEVEKYLKLDINSITDIYI	60
Qy	61	NTYKSGRNKALKKFKXYLTMEVLELKNNSLTPEKKNLHFVIGGQINDTAINYNQWKD	120
Db	61	DTYKSGRNKALKKFKXYLVTEVLELKNNLFPVEKNLHFVIGGQINDTAINYNQWKD	120
Qy	121	VNSDYTVKYFDNSAFNLTKTIVESATNTLTESFRNLNDPFDYKFKRMEIYY	180
Db	121	VNSDYVNVFYDNSAFNLTKTIVESAINDTLTESFRNLNDPFDYKFKRMEIYY	180
Qy	181	DKQKHFDIYKSGIENPEFIIDNIITKYLNSKDLALNKYIEESLNKNTANNNDI	240

Db	181	DKQXNFYNYKAQREENPELIIDIVKTYLSNEYSKEIDELNTYIEESLNKNTQNSGV	240
Qy	241	RNLKFADEDLVRLYNOELVERNLAASDILIRISMLKEDGGVYLDVLDLPGQPDLPKS	300
Db	241	RNFEPKNGESFNLYEOELVERNLAASDILIRISALKEIGGYLDVMDLPGQPDLPFS	300
Qy	301	INKPDSINTSWEMIKLEAMKKEYIPGYTSKNFMDLDEEVQSPFSALSSKSDSEIF	360
Db	301	IEKSSVTVDWFEMTKLEALWKKEYIPYTSBHFMDLDEEVQSPFSVLASKSDSEIF	360
Qy	361	LPLDDIKVSPLEVKIAFANNVINOALISLKSVCSDLVINOIKRYKILNDNLNPSINE	420
Db	361	SSLGMEASPLEVKIAFNKGIINOGLISVKDSYCSNLIVKQIENRYKILNLSNPAISE	420
Qy	421	GTDFNTWKIFSDKLASISNEDNMFMKITYNYLKVGFADPDRSTINLSGPGVVTGAYD	480
Db	421	DNDFNTTNTTIDSGIMAEANADNGRFYMLGKYLRVGFPPDVKTITINLSGPEAYAAQD	480
Qy	481	LLMPKDNSTNTHLPELENFEFPKTKISOLTEQETISLWSFNQARAKSQPEEYKGYFE	540
Db	481	LLMPKEGSMNIHLHEADLRNFEISKTNISQSTQTEGMASLWSPDDARAKAQPEEYKRYFE	540
Qy	541	GALGEDDNLDPQNTVLDKDYVSKKILSSMKTRKEYIHYIVLQGDKISYEASCNLPFSK	600
Db	541	GSLGEDDNLDPSONIVVDKEYLLEKISLSARSSRGYIHYIVLQGDKISYEAACNLPFAK	600
Qy	601	DPYSILYQKNIQEGSETAYYYVADAETKEIDKVRIPYQISNKNKIKLTFTHGKSEPT	660
Db	601	TPYDSVLFQKIEDSEIAYYVNPQDGEIQEIDKVKIPSIISDRPKIKLTFTHGKDEFT	660
Qy	661	DFANLVDVDSLSSEBETILNLAADISPKYIEINLLGCMFSPYSISABETYPGKLLKIK	720
Db	661	DIFAGFDVDSLSTEIEAADIKEDISPKSIEINLLGCMFSPYSINVEETYPGKLLKVK	720
Qy	721	DRVSELMPSISODSITVGSANQVEYVINEEGREIILDHSGKWNKEESIIOISSKEYISF	780
Db	721	DKISELMPSISODSIIIVGSANQVEYVINEEGREILDHSGEWINKESIIOISSKEYISF	780
Qy	781	NPKNKIIVKSXYLHELSTLQEIERNNANSDDILEKKVMLTECEINVASNIDROIVGR	840
Db	781	NPKNKIITVKSXNLPSTLLQEIERNNANSDDILEEKVMLTECEINVISNIDTQIVER	840
Qy	841	IEAKNLTSDSINYIKNEPKLIESDSLYLKHQGLDDSHFISFEDISKTEGFRFRF	900
Db	841	IEAKNLTSDSINYIKNEPKLIESDALCDLQKQNELEDSHFISFEDISETEGFSIRF	900
Qy	901	INKETGNSFIETEKIIFSEYATHISKETISNKTIDFDMVNGKLVKKNLDAHEVNTLN	960
Db	901	INKETGESIFVETEKITIFSEYANHTIEESKIKGIFDVMNGKLVKKNLDTTHEVNTLN	960
Qy	961	SAPFIQSLIEYNTTKESLSNLSVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDETID	1020
Db	961	AAPFIQSLIEYNSKESLSNLSVAMKVQVYAOQFSTGLNTITDAAKVVELVSTALDETID	1020
Qy	1021	LLPFTLSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQIEAKIGIMAVNLTAATAIYT	1080
Db	1021	LLPFTLSEGLPIIATIIDGVSLGAAIKELSETDPLLRQIEAKIGIMAVNLTAATAIIT	1080
Qy	1081	SALGIASGFSILLVPLAGISAGIPSVNNELLIOQKATKVIDYFKHISIAETEGFTLLD	1140
Db	1081	SSLGIASGFSILLVPLAGISAGIPSVNNELVLRDKATKVDFYFRHVSJVEVEGFTLLD	1140
Qy	1141	DKIIMPDDLVLTSEIDFNNSITFLGCEIWRABGGSGHTLTDIDHFFSSPSITYRKPMW	1200
Db	1141	DKIIMPDDLVITSEIDFNNSIVLGCETIWRMEGGSGHTVDDIDHFFSAPSTIYREPHL	1200
Qy	1201	SIYDVNLINKKEKIDFSGKDLMLFNAPNRFVGYEMGWTGPFERSLNDGTKLLDIRHYEG	1260
Db	1201	SIYDVLEVOKEELDLSKDLNVLNAPNRFVAMETGTGTLRSLENDGTKLLDIRDNYEG	1260
Qy	1261	QFYWRYFAFTADALITKLKPRYEDTNRINLDGNTSFIVPVITTEQIRKNLSYFVSGS	1320

Db 1261 EFYRYFAFADALITTLKPRVEDTNIRINLDSNTRSFIVPIITTEVIREKLSYFYGSG 1320
Qy 1321 GSYSLSPYMMIDLNVENDTWIDVDNVVKNITIESEDOKEGLIENILSKLNIEN 1380
Db 1321 GTVALSLQYMGINIELSESDVMIIDVDNVVEDVTIESDKIKGDLIEGILSTISEEN 1380
Qy 1381 KIILNHTINFGDINESNFIILSTFSLIEDINIIIEIDLVSXKILLSGCMKLIENS 1440
Db 1381 KIILNHEINFSEVNGSFGVSLTFSILLEGNAIEVLLLSKSYKLLISGELKILMNS 1440
Qy 1441 SDIQKIDHIFGFXGHEQKIPYISYIDNETKYNFIDYSKKEGLFTAEFNSIIRNIYMP 1500
Db 1441 NHIQKIDYIGFNSQLKNIPYFVDSGKENGFGSTKEGLFVSELDDVVLISKVYMD 1500
Qy 1501 DSNLFIYSKDKLDIRINKGDVKLLIGNYFKODMKVLSFTIETNTIKNGVLDEN 1560
Db 1501 DSKPSFGYGNLKVITKDNVNILTYLTKDDIKISLSLTLODEKTIKLSNVLDES 1560
Qy 1561 GVAQILKPMNAKASLNTSLSNLELESINIKIFYNLDPNLEFLDTNFIISGNSIG 1620
Db 1561 GVAELKPMNR-KGNTNTSLSMFLSMMIKSIFVNFQSNIKFILDANFIISGTSIG 1619
Qy 1621 QFELICDKXNIOPYFINFKIKETSYTYVGNRQNLIVEPSYHLDDSGNISSTVINFSOK 1680
Db 1620 QFELICDENDNIQYFIKFTLENTYTYVGNRQNLIVEPSYHLDDSGNISSTVINFSOK 1679
Qy 1681 YLXGIDRYNKKVIAPNLYT 1700
Db 1680 YLYGIDSCVKNWISPIY 1699

RESULT 4
US-10-011-366-6
Sequence 6, Application US/10011366
Publication No. US20030054493A1
GENERAL INFORMATION:
APPLICANT: Williams, James A.
Kink, John A.
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
DISEASE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/011.366
FILING DATE: 16-NOV-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/957.310
FILING DATE: 23-OCT-1997
APPLICATION NUMBER: US 08/329.154
FILING DATE: 24-OCT-1994
APPLICATION NUMBER: US 08/161.907
FILING DATE: 02-DEC-1993
APPLICATION NUMBER: US 07/985.321
FILING DATE: 04-DEC-1992
APPLICATION NUMBER: US 07/429.791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01121

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6
Query Match 51.5%; Score 4467; DB 14; Length 2710;
Best Local Similarity 50.7%; Pred. No. 9.8e-271;
Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;
Qy 1 NMLVKAQQLQWVYVFKRIQEDYVAINALBEEYHNMSESSVVEKYLLKLDINNLTNYL 60
Db 1 MSLISKEELIKLAY-SIRPRENEYKITLNLDEYKLTNNENKYLQKLKLNESIDVFM 59
Qy 61 NTYKSGRGNKALKKFKKYLWVLELKNLSLTPVEKNLHFIWIGQINDTAINIOWKD 120
Db 60 NKYTSSRNRLSNLKKDILKEVILIKNSTSPVEKNLHFVWIGVSDIALEYIKOWAD 119
Qy 121 VNSDYTKVYFVDSNAFLINTLKTIIVESATNNTLESFRENLDPEFDYKFKRMEIY 180
Db 120 INAEYNIKLYWYDEAFVLTUKAIVESSITTEAQLLEBEIQNFQDNMFKYKRMFIY 179
Qy 181 DKQHFIDYKSOIEENPEFTIDNIITKYLSEYKOLEALNKYIEESLNKITANNNDI 240
Db 180 DRQRFINYKSOINKPTVPTIDDIKSHLVSEYNRDETVELSYRTNSLRKXNSHGD 239
Qy 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVLDLPGIQDLPKS 300
Db 240 RANSLFTEQELNIVSQELNLRGNAASDIVRLALKNGVYLDVDMPLGHSDFPKT 299
Qy 301 INKPSITNTSWEMIKLEAIKMYKEYIPGYSKPKDMLDERVQSPSALSCKSKSEIF 360
Db 300 ISRPSSIGLDRWEMIKLEAIKMYKYNNTYSENFDKLQQLKONFKLIIESKSEIF 359
Qy 361 LPLDDIKVSPLEVIAPANNVINOALISLKDSCDLVINOIKRYKILNDNLNPSINE 420
Db 360 SKLENLVSDLEIKIAPALGVSINOALISKGSYLTNLVIEQVKNRQFNLQNLNPALES 419
Qy 421 GTDFNTTMKIPSDKLASINEDNMFMKITYNLKVGFPAPDVRSTINLSGCVTGYAD 480
Db 420 DNNFTDTTKIPIHDSLFNSATAENSFLTKIAPYIQVGFMPARSTISLGSFGASAYD 479
Qy 481 LLMFKONSTNIHLEPELNRNFEPPKTKISQTEQITSLMSFNQARAKQPEEYKGYFE 540
Db 480 FINLQNTIETKASDLIEFKFPENNLSQTEQINSLSFQASAKYQEKYVRDYG 539
Qy 541 GALGEDDNLDPQNTVLDKDY-VSKKILSS--MKTRNKEYIHYIVLOQDKIYEASCNL 597
Db 540 GSLSDNGVDNFKNTALDKNLLNKNIPSNVVEAGSKNYVHYIQLGGDISEATCNL 599
Qy 598 FSKOPYSIILYKNIQEGSETAYVYVAD--AEIKEDKYRIPYQISNKNIKLFIHGK 655
Db 600 FSKNPNSTIIQRNM--NESAKSYFLSDGSEIIEELNKRIPERLKKKVKYFIHGK 657
Qy 656 SEFTDTTANLDDVLSLSEIETILNLAADISPKYIEINLILGCMNFSYSISAEETPEKL 715
Db 658 DEFTSEFARLSVDSLSEISFLDTIKLIDISPKNVEVNLGCMNFSYDFNVEETPEKL 717
Qy 716 LKIKDRVSELMPSISQSDITVSANQVEVRINEGKREILDHSGKWINKESESIDKISSK 775
Db 718 LLSIMDKITSTLPDVNKNISITIGANQVEVRINSEGRKELLAHSGKWINKESEIMSDLSK 777
Qy 776 EYISFNPKENKLIYKSKYLHELSTLLOEIRNANSSDIDLEKVMLECEINVASNDRQ 835
Db 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLLDASVSPDTKFIILNKLNISSIGDY 837
Qy 836 IVEGRIEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDDSHFISFEDISKTEG 895

Db 540 GSLSENGVDNFKNATLADKNYLLANKIPSNVVEEAGSKNYHYIIQLQCDDISYEATCNL 599
Qy 598 FSKDPYSSILYQXNIGESTAYYYVAD--AEIKEDKYRIPIYQISNKENIKLTGIGHGK 655
Db 600 FSKNPKNSIIQORN--NESAKSYFLSDGSEILNKRYPERLKNKEKVKVTFIGHGK 657
Qy 656 SEFNTDTPANLDVDSLSSEIETILNKAADIGPKYIEINLGCNMFYSISAEETYPGKL 715
Db 658 DEFNTSEPARLSVDSLSNISISFLDTIKLIDSPKVEVNLGCNMFSDPVEETYPGKL 717
Qy 716 LKIKDRVSELMPSIQSDSITVSANQYEVRIINEBKGREILDHSGKWINKESIIKDISK 775
Db 718 LLSIMDKITSLPDVKNKSITIGANQYEVRIINEBKGREILDHSGKWINKESIIKDISK 777
Qy 776 EYISFNPKNKIIVSKYHEILSTLQIBRNANSSDIDLEKXVMTCEINVASNIDRQ 835
Db 778 EYIFFSDIDNKLKAKSNIPGLASISEDIKTLLDASVSPDKFTILNKLKNIESIGDY 837
Qy 836 IVEGRTEERAKNTSDSINIKNEFKLIESISLVDLKHQGLDSDSHFISPEDISKTEG 895
Db 838 IYKELEPVKNIIHNSIDIDLEFNLENVSDELYELKKNLDEKYLISFEDISKNST 897
Qy 896 FRIRFINKETGSIETETKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNVDAAE 955
Db 898 YSVRFINKNGESVYVETEKEIFSKYSEHTKEISTIKNSIITDVNGLDNIQLDHTSQ 957
Qy 956 VNTLNSAFFIQSLEIYNTTKESLNSVAKVQVLAQFSTGLNTITDASKVVELVSTAL 1015
Db 958 VNTLNAAFFIQSLEIDYSSNKNVDLSTSVQVLAQFSTGLNTIYDSIQVNLISNAV 1017
Qy 1016 DETIDLPLSEGLPIATIIDVSLGAAIKELSETNDPLRQETAKIGIVAVNLTAAS 1075
Db 1018 NDTINVLPITTEGIPVITLDGINLGAIKELDEHPLLKKELEAKGVIAINLSI 1077
Qy 1076 TAVTSALGTASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETGA 1135
Db 1078 AAVTASIVGAEVITFLPIAGISAGIPSLVNNELIHKATSVVYFNHLSBSKYG 1137
Qy 1136 FTLLDDKIIMPQDLVLSEIDFNNSITLKGCEIMRAEGSGHTLTDIDHFFSPSPITY 1195
Db 1138 LKTEDDKILVPIDDLVISEIDFNNSISIKLGTGNLAWEGSGHTVTGNIDHFFSPSPIS 1197
Qy 1196 RKPWLSIYDLNTEKBIKIPSKDLAVLPNAPRVFQYEMGWTPGFRSLNDGTGLDRIR 1255
Db 1198 HIPSLSIYSAIGIETENLDPFSKKIMLPNAPRVFWWETGAVPGLRSLNDGTGLDRIS 1257
Qy 1256 DHYEQPYWYPAFIADALITKLKPYEDTNVPIINDGNTRSPFIVPITTEIRKNLSYS 1315
Db 1258 DLYPGKPYWYPAFF-DYAITTLKPYEPTNIXIKLKDTRNFMPTITTNIRKNLSYS 1316
Qy 1316 FYGSGGSYLSLSPYNNIDNLVENDTWVIDVNVKNITTESDEIOKGELIENILSKL 1375
Db 1317 PDGAGGYSLLSYPLSTNINSKODLFIENIDNEVREISLNGTIRKGLIKOVLSKI 1376
Qy 1376 NIEDNKIILNNHINIFYGDINESNRFTISLTFISLEIDILVSKYSKILLSGNCKM 1435
Db 1377 DINKKLIIGNQIDFSGDIDNDRYIFLTCELDDKISLIEINLVAKYSYLLSGDKY 1436
Qy 1436 LIENSSDIQKIDHIGBHQKYPYSYID-NETKNGPIDYSKKEGLFTAEFNSIESII 1494
Db 1437 LISNLSNTEIKINTGLD--SKNIAVNTDESNNKYFGAI-----SKTSQKSII 1483
Qy 1495 RNITMPSNKL-----FIYSKOL--KDIRINKGDVKLLIGNYPXD---DMKVSLS 1541
Db 1484 H--YKDSKNILEYNDSTLEFNSKOFIABDINVMKODINTITKYYVNDNTDKSIDFS 1541
Qy 1542 FTIEDTNTIKLVYLDENGVAQILKFMNNAKALANTSNLSMNFLESINIKIFNNLDP 1601
Db 1542 ISLVSKNOVKNGLYLINESYSSVLDVFKVSDGHNTSNFNLFLDNISFWKLFGFE--- 1598
Qy 1602 NIEFILTNTIISGNSIGOFELICDKOKNIQPIFNFKISETSYLYVGNRQNLIVERPS 1661

Db 1599 NINPVIDKYFTLVGKTNLGVFEICDNNKNIDIFYGKWTSSSKSTIFSGNGRNVVPEPI 1658
Qy 1662 YHLDDSNISSTYSQKLYGIDRYVKNKVIAPNLYT 1700
Db 1659 YN-PDGTGEDISTSLDSYEPGLYGDYINKVLIAPDLYT 1696
RESULT 6
US-10-271-012-6
; Sequence 6, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-271-012-6
Query Match 51.5%; Score 4467; DB 15; Length 2710;
Best Local Similarity 50.7%; Pred. No. 9.8e-271;
Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;
Qy 1 MNLVKKQLQKQVVKVFRIOEDYVAILNALVEEYHNMSESVBKYKLKDINKLNTDYL 60
Db 1 MSLSKELIKLAY-SIRPRENEVKTILTNLDEYKLTNNENKYLQKLNESIDVFM 59
Qy 61 NTKYKSGNKKALKFKPKYLTWVELEKNSLTPVEKNLHFWIGQINDTAINQWKD 120
Db 60 NKYKTSSENRLSKKDLKEVILIKNSNTPSPVEKNLHFWIGVIGVSDIALEYIKQWAD 119
Qy 121 VNSDYTKVFFVDSNAFLINTLKKTVESATNNTLESFRENLDPEFDYKFKRMEIY 180
Db 120 INAEYTKLWYDSBAFLVNTLKAIVSSSTTEALQLLEEIQNFQDNMKFKRMEFIY 179
Qy 181 DKQHFIDYKQSQZEENPEFTIDNIITKLYSNEYSKDLALNKYIEESLNKITANNNGDI 240
Db 180 DRQKRFINYKSLQNKFTVPTIDDIILKSLVSEINRDETLYESVTSLSRLKNSHGDID 239
Qy 241 RNLEKFADELDVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDLPKS 300

Db 240 RANSLTEQELINISYQOELNLRGNAASDIIVALLAKNFGVYLDVMDLPGHSDLFKT 299
Qy 301 INKPDSTINTSWEMIKLEALMKYKEVPGVTSKNFMDLBEVQORSFESALSXSXSKSEIF 360
Db 300 ISRPSSIGLDRWEMIKLEALMKYKINNYTSENFPKLDQQLKDNFKLIIESKSEKSEIF 359
Qy 361 LPDDDKVSPLEVKIAPANNVINOALISLKDSCVSDVAINQIKREYKILNDMLNPSINE 420
Db 360 SKLENVSDLEIXIAFALGSVINQALISKQGSYLTNLVIEQVKRYQFLNQHLPALIES 419
Qy 421 GTDFNTMKIFSDKLASINEDNMFMKILNTLVKGFAPDVRASTINLSGPGVYTGAYOD 480
Db 420 DNNFTDTTKLPHSLFNASAEHSMFLTKIAPYLVGFMPPEARSTISLSPGAYASAYD 479
Qy 481 LLAFKDNSTNHLLEPRLNFPKTKISQLTQETITSUWSFNQARAKSQFBEYKGYPE 540
Db 480 FINLOENTIEKTLKASDLIEFKPPENLSQLTQETINSUWSFQASAKQFKEYVRYDTG 539
Qy 541 GALGEDNLOFAQNTVLDKDY-VSKILSS--MKTRKEVYIHHVLOQDKISYEASCNL 597
Db 540 GSLSEDNGVDFNKNTALDKNYLLNNKIPSNVVEAGSKYVHYIIQLOGDDISYEATCNL 599
Qy 598 PSKDPYSILYQKIEGSEATAYVYVAD--AEIKEDKYRIPYQISNKRNIKLTFIGHGK 655
Db 600 PSKPNKNSIITQRNM--NESAKSYFLSDGSESILELNKRIPELKNKEVKVTFIGHGK 657
Qy 656 SEFNTDFANLDVDSLSETEITILAKADISPKYIEINLLGCNMFYSISABETYPGKL 715
Db 658 DEFNTSFEARLSVDLSNEISSFLDTIKLIDISPKVNEVLLGCNMFYSYDNVEETYPGKL 717
Qy 716 LKIKDRVSELMPSIODSITVSANQYEVRAINEREGKEIILDHSGKWKINKEESIIOKISSK 775
Db 718 LLSIMDKITSTLPWNKNSITIGANQYEVRAINEREGKELLARHSGKWKINKEEAIMSLSSK 777
Qy 776 BYISFNPKENKIIVKSKYLHLSLTLLQEIERNANSDDIDLEKKVMTCECINVASNDRQ 835
Db 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLTLLDASVSPDTKPIILNNLKNIESIGDY 837
Qy 836 IVEGRIBEAKNLTSDSINVINKNEFKLIESDSYDLKHQNGLDDSHFISPEISIKTEG 895
Db 838 IYEEKLEVPKNIIHNSIDDLDEFNLLENVDELYELKUNLDEKYLISFEISKNST 897
Qy 896 FRIRPINKETGNSPIETKEIFSEYATHISKEISNIDKOTFDNVNGKLVKNVNDAAHE 955
Db 898 YSVRFINKNGESVYVETKEIFSKYSEHITKEISTIKNSIITDVNGNLLDNIQLDHTSQ 957
Qy 956 VNTLNSAFFIOGLIENYTKESLSNLSVMKQVYQALFSTGLNTITDASKVVELVSTAL 1015
Db 958 VNTLNAAPFIQSLIDYSNKNVDLSTSVKQVLAQLFSTGLNTIYDSIQLVNLSNAV 1017
Qy 1016 DETIDLLPTLSGLPIIATIIDGVSIGAAIKELSETNDPILROEIAKIGIMAVNLTAAS 1075
Db 1018 NDTINVLPTITEGIVSTILDGINLGAALKELDEHDPDLUKKELEAKGVIAINMBSLI 1077
Qy 1076 TAVTSALCIAGSGFILLVPLAGISAGIPSLVNNELIILQDKATKVIDYFKHISLAETEGA 1135
Db 1078 RAATVASIVIGAEVTFILPFIAGISAGIPSLVNNELIILHDKATSVNRYFNLSSEKXYP 1137
Qy 1136 FTLLDDKIMPODDVLVSEIDFNNSNITLKGCEIWRAGGGSHLTDDIDHFFSPSPITY 1195
Db 1138 LKXTEDDKILVPDVLVSEIDFNNSNIXLGTGTCNILAMEGGSGHTVTGNIDHFFSPSPISS 1197
Qy 1196 RKPWLSIYVNLNIKEKIDFSDKLVLPNAPNRVFGYEWGTPGPRSLDNDGTLLDORIR 1255
Db 1198 HIPSLSIYGAIGIETENLDFSKIMMLPNAPSRVFWETGAVPGRLSLENDGTLLDORIR 1257
Qy 1256 DHYEGOFYWRYPAFIADALITIKLKRYEDTNVIRINLDGNTRSFIVPVIITTEQIRKNLSYS 1315
Db 1258 DLYPGKFRWRFAFF--DYAITLKVPVEDTNIKIKLDKOTRNFNFTTITTEIRKNLSYS 1316
Qy 1316 FYGGSGYSLSISPPYNNIDMLVENDTWVIDVNVKNITTESDEIQKGELIENILSKL 1375

Db 1317 FDGAGGYTSLLSYPISTWINLSKODLMIFNIDNEVREISIENTGKIKGLIKDLVLSKI 1376
Qy 1376 NIEDNKIILNNHINIFYGDINESNRFTSLTFSILEDINIIIEIDLVSXSKYKILLSGNCKM 1435
Db 1377 DINKKLIIGNOTIDFSGDIDKNDRIFLTCEDDKISLIIIEINLVAKSYSLLSGDKNY 1436
Qy 1436 LIENSIOQKIDHIGPNGEHQKIPYSYD-NETKNGFIDYSKKEGLETFEAFSNEISII 1494
Db 1437 LISLSNTIEKINTGLD--SKNIAYNTEDESNNKYFGAI-----SKTSQKSI 1483
Qy 1495 RNITMPSNNL-----FIYSKDL--KDIRINKGDVKKLLIGNYFKD---DMKVSLS 1541
Db 1484 H--YKDSKNILEFYNDSTLEFNSKOFIABEDINVMKODINTITGKYVDNNTDKSDFDS 1541
Qy 1542 FTIEDTNTIKLVYLDENGVAQILKFMNNAKALANTNSLMMFLIESINIKNIFNNLDP 1601
Db 1542 ISLVSKQKVKNGLYNESVYSSVLDVFKNSDGHNTSNFNNLFLDNISFWKLFGEF-- 1598
Qy 1602 NIEFILTNTIISGNSIGOFELICDKDKNIQPIYFNFKIKETSITLYVGNRQNLIVEPS 1661
Db 1599 NINFVIDKYFLVGKTNLGVYEFICDNNKNIDIVFGWKTSKSSKSTIFSGNGRNVVVEPI 1658
Qy 1662 YHLDGSGNISSTVNFQSKYLGIDRYNVKVIAPNLYT 1700
Db 1659 YN-PDTGSDISTLSDFSVEPLYGIDRYNVKVIAPDDYT 1696

RESULT 7
US-10-463-957-1
; Sequence 1, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyles, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 556
; TYPE: PRF
; ORGANISM: Clostridium difficile
US-10-463-957-1

Query Match 25.4%; Score 2204; DB 12; Length 556;
Best Local Similarity 75.0%; Pred. No. 8.5e-130;
Matches 417; Conservative 66; Mismatches 73; Indels 0; Gaps 0;

Qy 1 MNLVNAQIQKMYVVKFRIQDEYVAILNALDEYHNMSSESVVEKYKLKDINNLTNDYL 60
Db 1 MSLVNRKQLEKMANVRFQDEYVAILDALSEYHNMSSESVVEKYKLKDINNLTNDYL 60
Qy 61 NTYKSGRNKALKKPKXYLTMEVLEKNSLTPVEKNLHFTWIGGINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKPKXYLTVEVLEKNNLTPVEKNLHFTWIGGINDTAINYNQWKD 120
Qy 121 VNSDYTVKYFSPNAPLINTLTKTIVESATNTLSPRENLDNPFDDYKNFKRWEIY 180
Db 121 VNSDYNNVYFSPNAPLINTLTKTIVESAINDTLSPRENLDNPFDDYKNFKRWEIY 180
Qy 181 DKQKHFDYKSOIEBPEFIIDNIKTLYSNEYSKDLEALNKYIEESLNKITAANGNDI 240
Db 181 DKQKNINYKAGREENPELLIDDIKTYLSNEYSKEIDELATYIEESLNKITAANGNDV 240
Qy 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVLDILPGIOPDLFKS 300
Db 241 RNFEFKNGSGNLYEQELVERWNLAASDILRISALKEIGGMVLDVMDLPGIOPDLFES 300
Qy 301 INKPDSTINTSWEMIKLEALMKYKEVPGVTSKNFMDLBEVQORSFESALSXSXSKSEIF 360


```

Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPEYTSSEHFDMLDEEVQSFSVLASKSKSEIF 360
Qy 361 LPDDIKVSPLEVKIAFANNVINOALISLKDYSYCSDLVINOIKRYKILNDNLNPSINE 420
Db 361 SSLGDWEASPLEVKIAFNKGIINOGLISVKUSYXSNLVKQLENYKILNNSLPAISE 420
Qy 421 GTDFNTMTKIFSDKLASINEDNMFMKITYLVKVGFPADVRSTINLSGPGVYTGAYOD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMMELGKYLVRGFPDPVKTINLSGPEAYAAAYOD 480
Qy 481 LLMPKDNSTNIHLEPELRNFPFKTKISQLTQEITSLWSFNQARAKSQFEYKKGYPF 540
Db 481 LLMPKEGSMNIHLIEADLRNFEISKTNISQTEQEMASLWSFDDARAKAQFEYKKNYFE 540
Qy 541 GALGEDDNLDPQNTV 556
Db 541 GSLGEDDNLDFSQNV 556

RESULT 9
US-10-463-957-12
; Sequence 12, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyles, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463.957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Clostridium difficile
; US-10-463-957-12

Query Match 25.2%; Score 2190; DB 12; Length 556;
Best Local Similarity 74.8%; Pred. No. 6.5e-129;
Matches 416; Conservative 66; Mismatches 74; Indels 0; Gaps 0;

Qy 1 MNLVNAQLQKQWVYKFRIOEDYVAILNALAEYHNMSSESVVEKYLKLDINNLTNYL 60
Db 1 MSLVNRKQLERKXANVRPRTQDEYVAILDALAEYHNMSNTVVEKYLKLDINSLTDIYI 60
Qy 61 NTYKSGRNKALKKPKYLTWEVLELKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKPKYLTVEVLELKNNSLTPVEKNLHFWAIGQINDTAINYNQWKD 120
Qy 121 VNSDVTVKVYDSNAFLNTLTKTIVESATNTLESFRENLDNDFPNKFKRMEIYY 180
Db 121 VNSDYNVNVFYDSNAFLNTLTKTIVESAINDTLESFRENLDNDFPNKFKRMEIYY 180
Qy 181 DKQKHFIDYKSOIBENPEFIIDNIITKYLSEYKOLEALNKYIEESLNKITTANNNDI 240
Db 181 DKQKFNFINYKQARENPELIIDDIIVKYLSEYKOLEALNKYIEESLNKITTQSGNDV 240
Qy 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFKNGESPNLYEQELVERWNLAASDILRISALKKEIGMYLDVMDLPGIQDPLFES 300
Qy 301 INKPSITNTSWEMTKLEAIMKYKEYIPGYTSKNFDMDEEVQSFSALSKSKSEIF 360
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPEYTSSEHFDMLDEEVQSFSVLASKSKSEIF 360
Qy 361 LPDDIKVSPLEVKIAFANNVINOALISLKDYSYCSDLVINOIKRYKILNDNLNPSINE 420
Db 361 SSLGDWEASPLEVKIAFNKGIINOGLISVKUSYXSNLVKQLENYKILNNSLPAISE 420
```

```

Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPEYTSSEHFDMLDEEVQSFSVLASKSKSEIF 360
Qy 361 LPDDIKVSPLEVKIAFANNVINOALISLKDYSYCSDLVINOIKRYKILNDNLNPSINE 420
Db 361 SSLGDWEASPLEVKIAFNKGIINOGLISVKUSYXSNLVKQLENYKILNNSLPAISE 420
Qy 421 GTDFNTMTKIFSDKLASINEDNMFMKITYLVKVGFPADVRSTINLSGPGVYTGAYOD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMMELGKYLVRGFPDPVKTINLSGPEAYAAAYOD 480
Qy 481 LLMPKDNSTNIHLEPELRNFPFKTKISQLTQEITSLWSFNQARAKSQFEYKKGYPF 540
Db 481 LLMPKEGSMNIHLIEADLRNFEISKTNISQTEQEMASLWSFDDARAKAQFEYKKNYFE 540
Qy 541 GALGEDDNLDPQNTV 556
Db 541 GSLGEDDNLDFSQNV 556

RESULT 8
US-10-463-957-11
; Sequence 11, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyles, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463.957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Completely synthesized
; NAME/KEY: MISC FEATURE
; LOCATION: (395)-(395)
; OTHER INFORMATION: Xaa at position 395 is ala, asp, glu, phe, gly, his, ile, lys,
; OTHER INFORMATION: leu, met, asn, pro, gln, arg, ser, thr, val, trp, or tyr.
; US-10-463-957-11

Query Match 25.3%; Score 2193; DB 12; Length 556;
Best Local Similarity 74.8%; Pred. No. 4.2e-129;
Matches 416; Conservative 66; Mismatches 74; Indels 0; Gaps 0;

Qy 1 MNLVNAQLQKQWVYKFRIOEDYVAILNALAEYHNMSSESVVEKYLKLDINNLTNYL 60
Db 1 MSLVNRKQLERKXANVRPRTQDEYVAILDALAEYHNMSNTVVEKYLKLDINSLTDIYI 60
Qy 61 NTYKSGRNKALKKPKYLTWEVLELKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKPKYLTVEVLELKNNSLTPVEKNLHFWAIGQINDTAINYNQWKD 120
Qy 121 VNSDVTVKVYDSNAFLNTLTKTIVESATNTLESFRENLDNDFPNKFKRMEIYY 180
Db 121 VNSDYNVNVFYDSNAFLNTLTKTIVESAINDTLESFRENLDNDFPNKFKRMEIYY 180
Qy 181 DKQKHFIDYKSOIBENPEFIIDNIITKYLSEYKOLEALNKYIEESLNKITTANNNDI 240
Db 181 DKQKFNFINYKQARENPELIIDDIIVKYLSEYKOLEALNKYIEESLNKITTQSGNDV 240
Qy 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFKNGESPNLYEQELVERWNLAASDILRISALKKEIGMYLDVMDLPGIQDPLFES 300
Qy 301 INKPSITNTSWEMTKLEAIMKYKEYIPGYTSKNFDMDEEVQSFSALSKSKSEIF 360
```


QY 421 GTDFTMTWKIFSDKIASISNEDNMFMKIITNLYKVGFPDVRSTINLSGPGVITGAYQD 480
DB 421 DNDFTNTTIFIDSIMAEANADGRFMWELGKLVGVFFEDVKTITNLSGPEAYAAAYQD 480
QY 481 LLMFKNSTNHLLEPELRFNPEPKTKISQITQETISLMSFNQARAKSQFEEYKGYE 540
DB 481 LLMFKEGSMNHLLEADLRNFEISKTNISQSTQEMASLWAFDDAKAQFEEYKRYE 540
QY 541 GALGEDNLDPAQNTV 556
DB 541 GSLGEDNLDPSQNTV 556
RESULT 10
US-10-463-957-9
; Sequence 9, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyrès, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463.957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 522
; TYPE: PRN
; ORGANISM: Clostridium difficile
US-10-463-957-9
Query Match 23.6%; Score 2051.5; DB 12; Length 522;
Best Local Similarity 73.8%; Pred. No. 2.9e-120; Indels 1; Gaps 1;
Matches 385; Conservative 66; Mismatches 70;
QY 35 HNMSESSVVEKYLKLDINLTDNTYNTYKSGRNKALKKFKYLTMEVLELKNNSLTPV 94
DB 1 HNMSENTVVEKYLKLDINSLTDYIDYIKSGRNKALKKFKYLTMEVLELKNNSLTPV 60
QY 95 EKNLHFIWGGQINTAINYNQKDVNSDYTVKVFYDSNAPLNTLTKTTIVESATNNTL 154
DB 61 EKNLHFIWGGQINTAINYNQKDVNSDYTVKVFYDSNAPLNTLTKTTIVESASNDTL 120
QY 155 ESFRENLDPEFDYKFKYKMEIIVDKQKPIDYKQIENPEFIIDNIKIYLSNEY 214
DB 121 ESFRENLDPEFDYKFKYKMEIIVDKQKPIDYKQIENPEFIIDNIKIYLSNEY 180
QY 215 SKDLEALNKYIEESLNKIFANNNDIRNLEKFADEDLVLNQLVERWNLAAASDILRI 274
DB 181 SKDIDELNAYIEESLNKVTENSGNDVRNPEEPTGTEVNLVEQLVERWNLAGASDILRV 240
QY 275 SMLKEDGGVLDVLPQIOPDLFKSINKPDSI-TNIGWMIKLEAIMKYEYIPGYTSK 333
DB 241 AILKNIGGVLDVDMPLGHPDLFKDINKPDSVKTAVDWEENQLEAIMKHYEYIPGYTSK 300
QY 334 NFMDLDEEVQSFESALSOKSEIFLPLDKIUSPLEVKIAFANNVINQALISLKDS 393
DB 301 HFTDLEEVQSFESVLSKSKSEIFLPLGIEVSPLVEVKIAFAGSIINQALISAKDS 360
QY 394 YCSDIATNINQIKRYKILNLPNSINEGTDFTNKKIPSKLASISNEDNMFMKITNY 453
DB 361 YCSDILIKIQIKRYKILNLTDLGPIISQGNDFNTNNGFESLGAANEISFIKIGSY 420
QY 454 LKYGFPADVRSTINLSGPGVITGAYQDLLMFKNSTNHLLEPELRFNPEPKTKISQITE 513
DB 421 LRVGFPEANTTVLSGPIYAGAYKDLJTFKEMS:DTSLSSSELNPEFPKVNISQATE 480
QY 514 QETISLMSFNQARAKSQFEEYKGYEGALGEDDNLDPQNT 555
DB 481 QEKNSLWQFNEERAKQFEEYKKNYEGALGEDDNLDPQNT 522

RESULT 11
US-10-463-957-3
; Sequence 3, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyrès, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463.957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 500
; TYPE: PRN
; ORGANISM: Clostridium difficile
US-10-463-957-3
Query Match 22.8%; Score 1979; DB 12; Length 500;
Best Local Similarity 75.0%; Pred. No. 9.5e-116;
Matches 375; Conservative 60; Mismatches 65; Indels 0; Gaps 0;
QY 1 MNLVNAQLOQVMYVKFESIQDEYVAIILNALLEEYHNMSESSVVEKYLKLDINLNTDYL 60
DB 1 MSLVNRKQLEKRVANVRFRQTQDEYVAIILDALEEYHNMSENTVVEKYLKLDINSLDIYI 60
QY 61 NTYKSGRNKALKKFKYLTMEVLELKNNSLTPVLEKLFHFIWGGQINTAINYNQKND 120
DB 61 DYKSGRNKALKKFKYLTVEVLELKNNSLTPVLEKLFHFIWGGQINTAINYNQKND 120
QY 121 VNSDYTVKVFYDSNAPLNTLTKTTIVESATNNTLSPFRENLDPEFDYKFKYKMEIY 180
DB 121 VNSDYNNVYFDSNAPLNTLTKTTIVESAINTLSPFRENLDPEFDYKFKYKMEIY 180
QY 181 DKQKPIDYKQIENPEFIIDNIKIYLSNEYSKDLEALNKYIEESLNKIFANNNDI 240
DB 181 DKQKFNITYKQREENPELIIDDIIVKTYLSNEYSKDELNTYIEESLNKIFANNNDI 240
QY 241 RNLEKFADEDLVLNQLVERWNLAAASDILRI-SMLKEDGGVLDVLPQIOPDLFKS 300
DB 241 RNFEFNGESFNLVEQLVERWNLAAASDILRI-SALKEIGMYLDVDMPLGIPQDLFES 300
QY 301 INKPDSTNTSWMEMIKLEAIMKYEYIPGYTSKFNMDLDEEVQSFESALSOKSEIF 360
DB 301 IEPSSVTVDVFWMTKLEAIMKYEYIPGYTSKFNMDLDEEVQSFESVLSKSKSEIF 360
QY 361 LPDLDIKVSPLEVKIAFANNVINQALISLKDSYCSDLVINQIKRYKILNLPNSINE 420
DB 361 SSLGDMASPLEVKIAFNKSGIINQGLISVKDSYCSNLIVQIKRYKILNLPNSINFAISE 420
QY 421 GTDFTMTWKIFSDKIASISNEDNMFMKIITNLYKVGFPDVRSTINLSGPGVITGAYQD 480
DB 421 DNDFTNTTIFIDSIMAEANADGRFMWELGKLVGVFFEDVKTITNLSGPEAYAAAYQD 480
QY 481 LLMFKNSTNHLLEPELRFN 500
DB 481 LLMFKEGSMNHLLEADLRN 500
RESULT 12
US-10-463-957-5
; Sequence 5, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyrès, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:37:55 ; Search time 63.75 Seconds

(without alignments)
8413.819 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700

Perfect score: 8675

Sequence: 1 MNLVNAQLQKVVYKFRIQ.....YLYGIDRYVKNVYIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTEMBL 25:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_rhnc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8675	100.0	2364	2 Q46342	Q46342 clostridium
2	6787	78.2	2367	2 Q9EXR0	Q9EXR0 clostridium
3	6758	77.9	2367	2 Q9F931	Q9F931 clostridium
4	6737	77.7	2367	2 Q46034	Q46034 clostridium
5	2156.5	24.9	554	2 Q93L39	Q93L39 clostridium
6	2082.5	24.0	2178	2 Q46149	Q46149 clostridium
7	1916	22.1	698	2 Q68553	Q68553 clostridium
8	1898	21.9	698	2 Q9EXQ8	Q9EXQ8 clostridium
9	1892	21.8	697	2 Q86141	Q86141 clostridium
10	1474	17.0	553	2 Q9FCX5	Q9FCX5 clostridium
11	512.5	5.9	3169	2 Q9ZGR4	Q9ZGR4 escherichia
12	507.5	5.9	3169	2 Q82316	Q82316 escherichia
13	494	5.7	3317	16 Q8SWP8	Q8SWP8 mycoplasma
14	489	5.6	2792	5 Q814R2	Q814R2 plasmodium
15	479	5.5	2771	5 Q26216	Q26216 plasmodium
16	468.5	5.4	2965	5 Q8WRS5	Q8WRS5 plasmodium

17 467 5.4 2752 5 Q9BJY0 Q9bjy0 plasmodium
18 461 5.3 2976 5 Q8WRS4 Q8wrs4 plasmodium
19 460 5.3 5767 5 Q81525 Q8i525 plasmodium
20 454.5 5.2 2957 5 Q8WRS6 Q8wrs6 plasmodium
21 448.5 5.2 3081 5 Q7YWF1 Q7ywf1 plasmodium
22 448.5 5.2 3203 5 Q7YWE7 Q7ywe7 plasmodium
23 448 5.2 3130 5 Q9BK46 Q9bk46 plasmodium
24 448 5.2 3130 5 Q8IDX6 Q8idx6 plasmodium
25 448 5.2 3254 5 Q9BK45 Q9bk45 plasmodium
26 448 5.2 3254 5 Q7YWE9 Q7ywe9 plasmodium
27 447.5 5.2 3080 5 Q7YWF0 Q7ywf0 plasmodium
28 447.5 5.2 3203 5 Q7YWE6 Q7ywe6 plasmodium
29 447 5.2 3096 5 Q7YWF2 Q7ywf2 plasmodium
30 447 5.2 3256 5 Q7YWE8 Q7ywe8 plasmodium
31 440 5.1 2269 5 Q26223 Q26223 plasmodium
32 439.5 5.1 2770 5 Q7YUE9 Q7yue9 plasmodium
33 436.5 5.0 2184 5 Q8IUQ6 Q8iuiq6 plasmodium
34 435.5 5.0 2747 5 Q9BJX9 Q9bjx9 plasmodium
35 433.5 5.0 2849 5 Q8IHV4 Q8ihv4 plasmodium
36 431.5 5.0 4688 16 Q9PC08 Q9pc08 ureaplasma
37 422 4.9 2166 16 Q51465 Q51465 borrelia bu
38 418 4.8 3381 5 Q8I2V4 Q8i2v4 plasmodium
39 412.5 4.8 1956 5 Q8IIE1 Q8iie1 plasmodium
40 406.5 4.7 6761 5 Q8IC77 Q8ic77 plasmodium
41 402 4.6 1685 5 Q8IRP9 Q8irp9 plasmodium
42 401.5 4.6 3223 2 Q8L1L9 Q8l1l9 escherichia
43 399.5 4.6 1711 5 Q8MWP2 Q8mwp2 plasmodium
44 399.5 4.6 3223 2 Q9RPH1 Q9rph1 escherichia
45 399 4.6 3724 5 Q77320 Q77320 plasmodium

ALIGNMENTS

RESULT 1

Q46342 PRELIMINARY; PRT; 2364 AA.
ID Q46342
AC Q46342;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Cytotoxin L.
OS Clostridium sordellii.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1505;
RN [1]_
RC STRAIN=6018;
RX MEDLINE=95369733; PubMed=7642137;
RA Green G.A., Schue V., Montell H.;
RT "Cloning and characterization of the cytotoxin L-encoding gene of
RT Clostridium sordellii: homology with Clostridium difficile cytotoxin
RT B.";
RL Gene 161:57-61(1995).
RN [2]
RC STRAIN=6018;
RX MEDLINE=96149194; PubMed=8544213;
RA Green G.A., Schue V., Girardot R., Montell H.;
RT "Characterisation of an enterotoxin-negative, cytotoxin-positive
RT strain of Clostridium sordellii.";
RL J. Med. Microbiol. 44:60-64(1996).
DR EMBL; X82638, CAA57959.1; -;
DR FIR; I40884; I40884.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR GO; GO:0006413; P:translational initiation; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR007577; Gly_transf_sug.
DR InterPro; IPR001950; TUF_SUI1.
DR Pfam; PF01473; CW_binding_1; 18.
DR Pfam; PF04488; Gly_transf_sug; 1.
DR PROSITE; PS01118; SUI1_1; 1.

SQ	SEQUENCE	2364 AA; 270576 MW; EAD8A4467A89BDBB CRC64;
	Query Match	100.0%; Score 8675; DB 2; Length 2364;
	Best Local Similarity	100.0%; Pred. No. 0;
	Matches 1700; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	MNLVNAQLQWVVKFRIQDEVAIINALAEVYHNMSESSVEKYLKLDIINNLTNYL 60
DB	1	MNLVNAQLQWVVKFRIQDEVAIINALAEVYHNMSESSVEKYLKLDIINNLTNYL 60
QY	61	NTYKSGRNKALKKFKKEYLTWEVLELKNLSLTPVEKNLHFIWIGGQINDTAIYNQWKD 120
DB	61	NTYKSGRNKALKKFKKEYLTWEVLELKNLSLTPVEKNLHFIWIGGQINDTAIYNQWKD 120
QY	121	VNSDYTVKVFVDSNAFLINTLKCTIVESAATNTLESFRENLDPEFDYKFKRMEIYY 180
DB	121	VNSDYTVKVFVDSNAFLINTLKCTIVESAATNTLESFRENLDPEFDYKFKRMEIYY 180
QY	181	DKQKHFDYKKSQJEENPEFIIDNI IKTYLSNEYSKDEALNKYIEBSLNKIITANNNGDI 240
DB	181	DKQKHFDYKKSQJEENPEFIIDNI IKTYLSNEYSKDEALNKYIEBSLNKIITANNNGDI 240
QY	241	RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQPOLFKS 300
DB	241	RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQPOLFKS 300
QY	301	INKPDSINTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEVOQSFESALSSKSDKSEIF 360
DB	301	INKPDSINTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEVOQSFESALSSKSDKSEIF 360
QY	361	LPLDDIKVSPLEVKIAPANNVINQALISLKDSYCSDLVINQIKRYKIINDNLNPSINE 420
DB	361	LPLDDIKVSPLEVKIAPANNVINQALISLKDSYCSDLVINQIKRYKIINDNLNPSINE 420
QY	421	GTDFNTTWKIFSDKLASTSNEDNMWMIKIINYLKVGFPADVRSTINLSGPGVYTGAYQD 480
DB	421	GTDFNTTWKIFSDKLASTSNEDNMWMIKIINYLKVGFPADVRSTINLSGPGVYTGAYQD 480
QY	481	LLMFKDNSTNHLLEPELNEFPFKTKI SOLTEQETISLMSFNQARAKSQFEEYKKGFFE 540
DB	481	LLMFKDNSTNHLLEPELNEFPFKTKI SOLTEQETISLMSFNQARAKSQFEEYKKGFFE 540
QY	541	GALGEDDNLDPQNTVLDKQVYKSKILSSMTRNKEYIHYIQLQGDKI SYEASCNLFSK 600
DB	541	GALGEDDNLDPQNTVLDKQVYKSKILSSMTRNKEYIHYIQLQGDKI SYEASCNLFSK 600
QY	601	DPYSILYQKNIEGSETAYYVYVADAETKEIDKYRI PYQISNKENIKLTITIGHKSEFNT 660
DB	601	DPYSILYQKNIEGSETAYYVYVADAETKEIDKYRI PYQISNKENIKLTITIGHKSEFNT 660
QY	661	DTFANLDVDSLSEIETTILNLAADISPKYTEINLLGCNMFYSISABETYPGKLLIKIK 720
DB	661	DTFANLDVDSLSEIETTILNLAADISPKYTEINLLGCNMFYSISABETYPGKLLIKIK 720
QY	721	DRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWINKEESI IKDISKEYISF 780
DB	721	DRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWINKEESI IKDISKEYISF 780
QY	781	NPKENKIIVKSKYLHELSTLLOEIRNANSSDIDLEKKVWLTECEINVASNIDRQIVEGR 840
DB	781	NPKENKIIVKSKYLHELSTLLOEIRNANSSDIDLEKKVWLTECEINVASNIDRQIVEGR 840
QY	841	IBEAKNLTSDSNYIKNFKPLIESISDSLYDLKHONGLDSDHFTSFEDISKTEGPRIRF 900
DB	841	IBEAKNLTSDSNYIKNFKPLIESISDSLYDLKHONGLDSDHFTSFEDISKTEGPRIRF 900
QY	901	INKETGNSIFITEKEIPESEYATHISKEISNIKOTIFDNVANGKLVKKNLDAAEVNTLN 960
DB	901	INKETGNSIFITEKEIPESEYATHISKEISNIKOTIFDNVANGKLVKKNLDAAEVNTLN 960
QY	961	SAFFTQSILYENTTKESLSNLSVAMKVQVYLAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB	961	SAFFTQSILYENTTKESLSNLSVAMKVQVYLAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 2

QY	Q9EXR0	PRELIMINARY; PRT; 2367 AA.
DB	Q9EXR0	PRELIMINARY; PRT; 2367 AA.
QY	AC	Q9EXR0; 01-MAR-2001 (Tremblrel. 16, Created)
DB	AC	Q9EXR0; 01-MAR-2001 (Tremblrel. 16, Last sequence update)
QY	DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)
DB	DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)
QY	DE	Toxin B.
DB	DE	Toxin B.
QY	GN	TCDB.
DB	GN	TCDB.
QY	OS	Clostridium difficile.
DB	OS	Clostridium difficile.
QY	OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
DB	OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
QY	OX	NCBI_TaxID=1496;
DB	OX	NCBI_TaxID=1496;
QY	RN	[1]
DB	RN	[1]
QY	RP	SEQUENCE FROM N.A.
DB	RP	SEQUENCE FROM N.A.
QY	RC	STRAIN=8864;
DB	RC	STRAIN=8864;
QY	RA	von Bichel-Streiber C.;
DB	RA	von Bichel-Streiber C.;
QY	RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DB	RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
QY	RN	[2]
DB	RN	[2]
QY	RP	SEQUENCE FROM N.A.
DB	RP	SEQUENCE FROM N.A.
QY	RC	STRAIN=8864;
DB	RC	STRAIN=8864;
QY	RA	Kohl M.;
DB	RA	Kohl M.;
QY	RT	"The pathogenicity locus of Clostridium difficile, strain 8864."
DB	RT	"The pathogenicity locus of Clostridium difficile, strain 8864."

Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.	
EMBL; AJ011301; CAC19891.1; -	
PIR; A27636; A27636.	
GO; GO:0003743; P:translation initiation factor activity; IEA.	
GO; GO:0006413; P:translational initiation; IEA.	
InterPro; IPR002479; C:binding.	
InterPro; IPR007577; G:transf_sug.	
InterPro; IPR001950; TIF_SUI1.	
Pfam; PF01473; C:binding 1; 16.	
Pfam; PF04488; G:transf_sug; 1.	
PROSITE; PS01118; SUI1_1; 1	
SEQUENCE 2367 AA; 269072 MW; BF7087C92E7388E7 CRC64;	
Query Match 78.2%; Score 6787; DB 2; Length 2367;	
Best Local Similarity 76.6%; Pred. No. 2.2e-249;	
Matches 1303; Conservative 201; Mismatches 195; Indels 2; Gaps 2;	
QY 1 MNLVKAQLOKVVYKFRQEDVEYVAILNALREYHNMSESSVVEYKYLKLDINNTDNYL 60	
DB 1 MSLVNRKQLEKMANVRFRQEDVEYVAILDALREYHNMSESSVVEYKYLKLDINNTDNYI 60	
QY 61 NTYKSGRNKALKKFKXEYLTMEVLELKNLSLPVEKNLHFIWIGQINDTAINYNQWKD 120	
DB 61 DTYKSGRNKALKKFKXEYLTMEVLELKNLSLPVEKNLHFIWIGQINDTAINYNQWKD 120	
QY 121 VNSDYTVKVFYSNAFLINTLKXIVESATNTLTSEFRENLDNDFDYNNKFKRMEIY 180	
DB 121 VNSDYNNVYFYSNAFLINTLKXIVESASNDLTSEFRENLDNDFBNHTAFPRKMQIY 180	
QY 181 DKQKHIDYKSOIENPFIDNIITKYLSEYKDLKALNKYTESINKITANGNDI 240	
DB 181 DKQONFINTYKAKENPDLIIDIVKYLSEYKDLIDELNAYTESLNKVTENGNDV 240	
QY 241 RNLEKAEDELVLNQLVELERNLAAADILRISMLKEDGGYLDVDPGIGQDLPFKS 300	
DB 241 RNKEEFTGEVFNLYQELVERNLGASDILRAVLKNGYLDVDMPLGHPDLPKD 300	
QY 301 INKPDST-TNTSWEIMKLEAIMKYKEYIPYITSKNFMDLDEEYQSFESALSSKDKSEI 359	
DB 301 INKPDVKTAVDWEQLEAIMKYKEYIPYITSKHFDTLDEEYQSFESVSLAKSKDKSEI 360	
QY 360 FLPLDDIKVSPLEVKAFANNSVINQALISLQSYCSDIVNIOIKRYKILNDNLNPSIN 419	
DB 361 FLPLGGIEVSPLEVKAFAGSIIDQALISAKSYCSDLLIKQIKRYKILNDTLGPIIS 420	
QY 420 EGTDFNTWKIPDKLASINEDNMFMKITYNLKVGAPDVRSTINISGGVYTGAYQ 479	
DB 421 QGNDFTNMNFGESLGAIAENEISFIAKIGSLRVGFYPEANTTITUSGPTIYAGAYK 480	
QY 480 DLLMFKDNSTNHLLEPELNFPPFKTKISQLTEQETLSLWSPNQAPAKSQPEEYKGYF 539	
DB 481 DLLTFKMSIDTSLSELNFPFPKVNISQATEQEKNSLWQNEERAKIQPEEYKGYF 540	
QY 540 EGALGEDNDLDFAGNTYLDKDYVSKYLSSMKTRNKEYIHYIYVQLQDKISYEASCNLPS 599	
DB 541 EGALGEDNDLDFSQNTVTDKYLELLEKISSSTKSSRGVHYIYVQLQDKISYEACNLEA 600	
QY 600 KDPYSSILYKQNIKRGSEFAYVYVADABEIKEDYRIPYQISKNRKLFTIGHGKSEFN 659	
DB 601 KNPYDSILFQNIKEDSEVAYYNPTDSEIQEIDKIRIPDRIPKIKLTLIGHGKAEFN 660	
QY 660 TDTFANLDDVLSSEIETILNLAJADISPKYIEINILGCMNFSYSABETYPKLLIKI 719	
DB 661 TDIIFAGLDVLSSEIETILDLAKADISPKSIEINILGCMNFSYSVNVETYPKLLLRV 720	
QY 720 KDRVSELMPSISQDSITVSANQYEVRIINEGKRILDHSGKWNKEESIIDKISSKEYIS 779	
DB 721 KDKVSELMPSISQDSIIVSANQYEVRIINSEGRRLLDHSGEWINKESIIDKISSKEYIS 780	
QY 780 FNPKNKIIVKSKYLHELSTLLQIRNANSSDIDLEKKVNLTECEINVASNIDRQIVEG 839	
DB 781 FNPKNKIIVKSKNLPESLTLQIRNANSSDIDLEKKVNLTECEINVASNIETQVVEE 840	

QY 840 RIEBAKNTSDSINIKNEFKLIBESIDSLYDLKHQNGLDSDHIFISPEDISKTNGPRIR 899	
DB 841 RIEBAKSLTSDSINIKNEFKLIBESIDSLYDLKHQNGLESHRIFISPEDISKTDEGFSIR 900	
QY 900 FINKETGNSFIETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKVNDAAHEVNTL 959	
DB 901 FIDKETGESIFVETEKAIFSEYANHITSEISKLDTIFDTVNGKLVKKVILDTAHEVNTL 960	
QY 960 NSAFFTQSLIENYTTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDSTI 1019	
DB 961 NAAFFTQSLIGNSKESLSNLVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDSTI 1020	
QY 1020 DLLPTLSEGLPIATIIDGVSLGAAIKELSETNDPLLRQIEHAKIGIMAVNLTAASTAIV 1079	
DB 1021 DLLPTLSEGLPIATIIDGVSLGASIKELSETSDPLLRQIEHAKIGIMAVNLTAATAII 1080	
QY 1080 TSALGIASGFSILLVPLAGISAGISPLVNNELILOQKATKVIDYKPHISLAPTEGAFILL 1139	
DB 1081 TSSLGIASGFSILLVPLAGISAGISPLVNNELILRAEAKNVVDYFGHISLAESEGAFILL 1140	
QY 1140 DDKIIMPQDDLVSEIDFNNSITIGKCEIWAEGSGHLLTDDIDHFPSSSIYTRKPW 1199	
DB 1141 DDKIIMPQDDLVSEIDFNNSITIGKCEIWRMEGSGHVTDDIDHFPSSSTTYREY 1200	
QY 1200 LSTYDVLNITKKEIKIDFSKDLVLPNAPNRVFGVWGWTGPRSLDNDDGKLLDRDRDHYE 1259	
DB 1201 LSTYDVLVDVKEELDKOLMVLNAPADRIFWERGWTGELASLENDGKLLDRDRDHYE 1260	
QY 1260 GQFWRMYFAFIADALITKLPRYEDTVNRINDGNTRSFIVPVITTEQIRKMSLSYFYGS 1319	
DB 1261 GQFWRFFAFIADSVITKLPRYEDTVNRISLDSNTRSFIVPVITTEYREKLSYFYGS 1320	
QY 1320 GGSYSLSLSPYNNNDINLVENDTWIDVNDVNVKNTIETSDBIQKELIENILSKLNIED 1379	
DB 1321 GGTVALSLSCYNNNINIENENDTWIDVNDVVRDVTIETSDKIKKGDLIENILSKLSIED 1380	
QY 1380 NKIILNHNTHFYGDINESNRFSILTSFIILEDINIIEIDLVSYSKILLSGCMKLIEN 1439	
DB 1381 NKIILDNHEINFSGTNGNGFVSLTSFIILEGINAVIEVDLLSKSVKILSELKTLMAN 1440	
QY 1440 SSDIQQKIDHIGFNGEHQKYPYISYIDNETKNGFYDYSKKEGLFTAEPNSIIRNIYM 1499	
DB 1441 SNSVQOKIDYIGLNSELOKNIPIYSFWDDDEKENGFINCTFKEGLFVSELSDVVLIIKVYM 1500	
QY 1500 PDSNNLFIYSKOLKDIRINKGDVKLLIGNYFKDKMKVSLSTIETDTWTIKLNGVYLDE 1559	
DB 1501 DNSKPPPGFYSNDLKDKVITTKDDVIIITGYILKDDIKISLSFTIQDKNTIKLNGVYLDE 1560	
QY 1560 NGVAQILKFWNNAKSALNTSNLWNELESINIKNIFVNNLDNPIERTLDTNFIISGNSI 1619	
DB 1561 NGVAEILKFMNK-KGSTNTSDLSMFLSNKISIFIKLSKNAKILLDTNFIISGTTFI 1619	
QY 1620 GQFELICDKQKNIOPYINFKIKETSVTLYVGNRQNLIVFPYSHLDDSGNISSTVNFQ 1679	
DB 1620 GQFELICDKQKNIOPYINFKIKETLTKYTLVGNRQNLIVFPYSHLDDSGNISSTVNFQ 1679	
QY 1680 KYLYGIDRYVNNKVIIPNLTYT 1700	
DB 1680 KYLYGIDSCVNNKVISPGIYT 1700	
RESULT 3	
Q9F931 PRELIMINARY; PRT; 2367 AA.	
ID Q9F931	
AC Q9F931	
DT 01-WAR-2001 (TRENBLrel. 16, Created)	
DT 01-WAR-2001 (TRENBLrel. 16, Last sequence update)	
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)	
DE Cytotoxin B.	
GN TCDB.	
OS Clostridium difficile.	
CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	
OC Clostridium.	

QY	840 RIEEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQNGLDDSHFISPEDISKTEGFSIR 899	
DB	841 RIEEAKSLTSDSINYIKNEFKLIESISDALYDLKQNELEESHFISPEDISKTEGFSIR 900	
QY	900 FINKETGNSIFITEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAAEHVNTL 959	
DB	901 FIDKETGESIFVETEKAFSEYANHITEEISKLDKDTIFDTVNGKLVKKVLTDAEHVNTL 960	
QY	960 NSAFTIQLSLEYNTKESISNLISVAMKVQVYQALFSTGLNTITDASKVVELYSTALDETI 1019	
DB	961 NAAFTIQLSLEYNTKESISNLISVAMKVQVYQALFSTGLNTITDAAKVVVELYSTALDETI 1020	
QY	1020 DLLPTLSEGLPIIATIDGVSLGAAIKELSETNDPLLRQIEIAKIGIMAVNLTAATAIV 1079	
DB	1021 DLLPTLSEGLPIVATIDGVSLGASIKELSETSDPLLRQIEIAKIGIMAVNLTAATAII 1080	
QY	1080 TSALGIASGFSILLVPLAGISAGISPLVNNELLQDKATKVIDYFKHISLAETEGAFILL 1139	
DB	1081 TSSLGIASGFSILLVPLAGISAGISPLVNNELLISAEAKVNDVYFGHISLAEGEAFILL 1140	
QY	1140 DDKIIMPQDDLVISEIDFNNSITLKGCEIWRAGSGGHTLTDIDHFPSSSITVYRKPW 1199	
DB	1141 DDKIIMPQDDLVISEIDFNNSITLKGCEIWRMEGSGHTVTDIDHFPSSASTVYREPY 1200	
QY	1200 LSIYDVNLITKKEIDFESKOLMVLNAPNRVFGYEMGWTGFRSLDNDGTLLDIRRDHYE 1259	
DB	1201 LSIYDVLDVKEBELDLSKOLMVLNAPDRIFGWERGWTGFRSLDNDGTLLDIRRDHYE 1260	
QY	1260 GQFYWRYFAFIADALITKLPRYEDTNVRINLGNTRSFIVPVITTEIQIRKLSYSFYGS 1319	
DB	1261 GQFYWRFFAFIADSVITKLPRYEDTNIRISLDSNTRSFIVPVITTEYREKLSYSFYGS 1320	
QY	1320 GGSYSLSLSPYNNMIDNLVENDTWIDVNVKNITTESDEIQKGELIENILSKLNIED 1379	
DB	1321 GGTVALSLSQYNNMIDNLVENDTWIDVNVVDRVTIESDKIKGGLIENILSKLSIED 1380	
QY	1380 NKIILNHTINFYGDINESNRFISLTFSEILEDINIIEIDLVSYSKILLSGNCMKLIEN 1439	
DB	1381 NKIILNHEINFSGLTNGNGFVSLTFSILEGINAVIEVDLLSKYSKVLISSELKTLMAN 1440	
QY	1440 SSDIQQKIDHIFGNGEHQKIPYISYIDNETKYNGFIDYSKKEGLFTAEPESNISIRNIYM 1499	
DB	1441 SNSVQKIDYIGLNSLQKNIPYSFMDDESGKENGFINCFKEGLFVSELSDVVLIIVKYM 1500	
QY	1500 PDSNNLFYSSKDLKDIRIINKGVDKLLIGNFYKDDMKVLSLFTIEDTNTIKLNGVYLDE 1559	
DB	1501 DNSKPPFGYYSNDLKQVKVITKDDVITITGYVLKDDIKISLSFTIQDKNTIKLNGVYLDE 1560	
QY	1560 NGVAQILKFNANAKSALNTSLSMNFLESINIKNIFYNNLDNPTIEFTLDTNFIISGNSI 1619	
DB	1561 NGVAEILKFMNK-KGSTNTSDSLSMFLSNVTKISFTIKLSNAKULIDTNFIISGTFPI 1619	
QY	1620 GQFELICDKNIOPIYFIRNEKIKETSYLYVGNRQNLIVEPSYHLDSDGNSISSTVINFSQ 1679	
DB	1620 GQFELICDKNIOPIYFIRNEKIKETSYLYVGNRQNLIVEPSYHLDSDGNSISSTVINFSQ 1679	
QY	1680 KYLYGIDRYVNNKVIAPNLYT 1700	
DB	1680 KYLYGIDSCVNVKVISPIYT 1700	
RESULT 3		
ID	Q9F931	PRELIMINARY; PRT; 2367 AA.
AC	Q9F931;	
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)	
DE	Cytotoxin B.	
GN	TCDB.	
OS	Clostridium difficile.	
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	
OC	Clostridium.	

NCBI_TaxID=1496;
[1]
RW SEQUENCE FROM N.A.
RC STRAIN=5340;
RX MEDLINE=20448897; PubMed=10992443;
RA Sambol S.P., Merigan M.M., Lively D., Gerdling D.N., Johnson S.;
RT "Toxin gene analysis of a variant strain of clostridium difficile that
causes human clinical disease";
RL Infect. Immun. 68:5480-5487(2000).
DR EMBL; AF217292; AAG18011.1; -.
DR PIR; A27636; A27636.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR GO; GO:0006413; P:translational initiation; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR001950; TIF_SUI1.
DR Pfam; PF01473; CW_binding_1; 17.
DR Pfam; PF04488; Gly_transf_sug; 1.
DR PROSITE; PS01118; SUI1_1; 1.
SQ SEQUENCE 2367 AA; 269337 MW; D5EE715ESBD41E2F CRC64;

Query Match 77.9%; Score 6758; DB 2; Length 2367;
Best Local Similarity 76.1%; Pred. No. 2.8e-248;
Matches 1295; Conservative 204; Mismatches 200; Indels 2; Gaps 2;

QY 1 MSLVNAQLQKVVYKRIQDEVAIINALEEYHNMSSESVKYLKLDINLTDNYL 60
DB 1 MSLVNRKQLEKXANVRFRVQDEVAIIDLALAEYHNMSSENTVVEKYLKLDINSLDTYI 60

QY 61 NTKYKSGRNKALKKFEYLTWEVLKLNLSLTPVEKNLHFWIGGOINDTAINYINQW 120
DB 61 DTYKSGRNKALKKFEYLTWEVLKLNLSLTPVEKNLHFWIGGOINDTAINYINQW 120

QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLTESFRENLDPEFDYKFRKMEIY 180
DB 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLTESFRENLDPEFDYKFRKMEIY 180

QY 181 DKQHFIDYKQSEIENPEFIIDNIIKTYLSNEYSKOLEALNKYIEBSLANKIPANNNDI 240
DB 181 DKQHFIDYKQSEIENPEFIIDNIIKTYLSNEYSKOLEALNKYIEBSLANKIPANNNDI 240

QY 241 RNLEKFADEDLVRLNQBLVERWNLAAASDLIRLSMLKEDGGVLDVLDIIPGIDPLFKS 300
DB 241 RNLEKFADEDLVRLNQBLVERWNLAAASDLIRLSMLKEDGGVLDVLDIIPGIDPLFKS 300

QY 301 INKPDSI-TNTSWEMIKLEATMKYKYEIPGYTSKNFMDLDEEVORSFESALSSKDKSEI 359
DB 301 INKPDSI-TNTSWEMIKLEATMKYKYEIPGYTSKNFMDLDEEVORSFESALSSKDKSEI 359

QY 360 FLPLDDIKVPLEVIAFANNSVINOALISLKDSCSLVINOIKRYKILNDNLNPSIN 419
DB 361 FLPLDDIKVPLEVIAFANNSVINOALISLKDSCSLVINOIKRYKILNDNLNPSIN 419

QY 420 EGTDRNTMKIPSDKLASISNDNNMFMKTNLYKVGAFDVRSTINLSPGVYTGAYQ 479
DB 421 QGNDENTMNNFGESIGALANENISFTAKIGSYLRVGFPEANTVTVLSGFTIYAGAYK 480

QY 480 DILLMKDSTNHLLEPELRNPEPKTKISQLTEQIITSLMSFNQARAKSOFEEYKGYF 539
DB 481 DILLMKDSTNHLLEPELRNPEPKTKISQLTEQIITSLMSFNQARAKSOFEEYKGYF 539

QY 540 EGALGEDDNLDAQNTVLDKDVSKILSSMKTRNKEYIHYIVLQGGKISYEASCNLFS 599
DB 541 EGALGEDDNLDAQNTVLDKDVSKILSSMKTRNKEYIHYIVLQGGKISYEASCNLFS 599

QY 600 KDPYSILYQKNIQEGSETAYYYYVADAETKIDKYRIPYQISNKNIKLTFIGHGKSEFN 659
DB 601 KNPYDILYQKNIQEGSETAYYYYVADAETKIDKYRIPYQISNKNIKLTFIGHGKSEFN 659

QY 660 TDTFANLVDVLSSEIETTLNKAADISPKYTEINLLGONMPSYSABEYTPGKLLKI 719
DB 661 TDTFANLVDVLSSEIETTLNKAADISPKYTEINLLGONMPSYSABEYTPGKLLKI 719

QY 720 KDRVSELMPSIQSDSITVSANQVEVRINEEGKEIILDHSGKWNKESIIKOISSKEYIS 779
DB 721 KDRVSELMPSIQSDSITVSANQVEVRINEEGKEIILDHSGKWNKESIIKOISSKEYIS 780

QY 780 FNPENKIIVKSKYLHLSLTLQIBRNNANSSDIDLEKKVMTCECHNANSDROAVEG 839
DB 781 FNPENKIIVKSKYLHLSLTLQIBRNNANSSDIDLEKKVMTCECHNANSDROAVEG 840

QY 840 RIEBAKNTSDSINYIKNEFKLIESISDSLYDLKHQGLDDSHFISPEDISKTEGFRIR 899
DB 841 RIEBAKNTSDSINYIKNEFKLIESISDSLYDLKHQGLDDSHFISPEDISKTEGFRIR 900

QY 900 FINKETGNSIETETEKEIFSEYATHISKEISNIDKTIIDNVNGKLVKKVNLDAAEVNTL 959
DB 901 FINKETGNSIETETEKEIFSEYATHISKEISNIDKTIIDNVNGKLVKKVNLDAAEVNTL 960

QY 960 NSAFFIQSLIENYTKESLSNLSVAMKVQVYVAQLFSTGLNTITDASKVVELVSTALDETI 1019
DB 961 NSAFFIQSLIENYTKESLSNLSVAMKVQVYVAQLFSTGLNTITDASKVVELVSTALDETI 1020

QY 1020 DLTPTLSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQIEAKIGIMAVNLTAASTAIV 1079
DB 1021 DLTPTLSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQIEAKIGIMAVNLTAASTAIV 1080

QY 1080 TSAIGIASGFSILLVPLAGISAGIPSLVNNELIILQDKATKVIDYFKHISLAETEGFTLL 1139
DB 1081 TSAIGIASGFSILLVPLAGISAGIPSLVNNELIILQDKATKVIDYFKHISLAETEGFTLL 1140

QY 1140 DDKITMPODDLVLSEIDFNNNISITLQKCEIWRAGEGSGHTLTDIDHPFSPSTYRKPM 1199
DB 1141 DDKITMPODDLVLSEIDFNNNISITLQKCEIWRAGEGSGHTLTDIDHPFSPSTYRKPM 1200

QY 1200 LSIYDVNLKKEKIDFSDKOLMVLNAPNRVFGYENGWTPGPRSLDNDGTLLDRIDRYE 1259
DB 1201 LSIYDVNLKKEKIDFSDKOLMVLNAPNRVFGYENGWTPGPRSLDNDGTLLDRIDRYE 1260

QY 1260 GQFYWEYFAFIADALITLKPVEYEDTVRINLDGNTSRFIVPVTTEOIRKNLSYSFYGS 1319
DB 1261 GQFYWEYFAFIADALITLKPVEYEDTVRINLDGNTSRFIVPVTTEOIRKNLSYSFYGS 1320

QY 1320 GGSYSLSLSPYNNMTDLNLVENDTVIDVDVNVKNITTESDEIOKGELIENILSKNIET 1379
DB 1321 GGSYSLSLSPYNNMTDLNLVENDTVIDVDVNVKNITTESDEIOKGELIENILSKNIET 1380

QY 1380 NKIILNNHTINFGYNESNRPISLTFSEIDENIIEIDLVSKSYKILLSGCMKLIEN 1439
DB 1381 NKIILNNHTINFGYNESNRPISLTFSEIDENIIEIDLVSKSYKILLSGCMKLIEN 1440

QY 1440 SSDIOQKIDHIGFNGEHQKYPYSDIDNETKXNGEIDYKKEGLPTAEFSNESIIRNTYM 1499
DB 1441 SSDIOQKIDHIGFNGEHQKYPYSDIDNETKXNGEIDYKKEGLPTAEFSNESIIRNTYM 1500

QY 1500 PDSNNLIYSSKDLKDIRIINKGDVLLIKNYFKDDMKVLSFTIEDTNTIKNGVYLDE 1559
DB 1501 PDSNNLIYSSKDLKDIRIINKGDVLLIKNYFKDDMKVLSFTIEDTNTIKNGVYLDE 1560

QY 1560 NGVAQILKEMNNAKALNTSNLSMNFLESINIKNIFYNNLDNPIEFLDTNFIISGNSI 1619
DB 1561 NGVAQILKEMNNAKALNTSNLSMNFLESINIKNIFYNNLDNPIEFLDTNFIISGNSI 1619

QY 1620 GQPELLCDKDKNIQYFINKFKTSYTLVGNRQNLIVPEPSYHLLDDSGNISSTVINFSQ 1679
DB 1620 GQPELLCDKDKNIQYFINKFKTSYTLVGNRQNLIVPEPSYHLLDDSGNISSTVINFSQ 1679

QY 1680 KYLYGIDRVYKVIIPNLTYT 1700
DB 1680 KYLYGIDRVYKVIIPNLTYT 1700

RESULT 4
Q46034
ID Q46034
AC Q46034;

PRELIMINARY; PRT; 2367 AA.

Qy	1680	KYLIGIDRYVKNVLIAPNLYT	1700	
Db	1680	KYLIGIDSCVANKWISPNYT	1700	
RESULT 5				
Qy	93139	PRELIMINARY;	PRT;	554 AA.
ID	Q93L39	AC	Q93L39;	
DT	01-DEC-2001	(TREMBLrel. 19, Created)		
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)		
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)		
DE	TcDB-C34	Cluster1-2 (Fragment)		
DE	TcDB-C34.			
OS	Clostridium difficile.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1496;			
RN	[1]	SEQUENCE FROM N.A.		
RP	STRAIN=C34.			
RC	MEDLINE=2123121; PubMed=11430410;			
RX	Mehlig M., Moos M., Braun V., Kalt B., Mahony D.B.,			
RA	von Eichel-Streiber C.;			
RT	"Variant toxin B and a functional toxin A produced by Clostridium			
RT	difficile C34."			
RT	FEMS Microbiol. Lett. 198;171-176 (2001).			
DR	EMBL; AJ294944; CAC1640.1; -.			
DR	InterPro; IPR007577; Gly transf. sug.			
DR	Pfam; PF04488; Gly transf. sug; 1.			
FT	NON TER 554			
SQ	SEQUENCE 554 AA; 63910 MW; 9285512AC1B1D87A CRC64;			
Query Match 24.9%; Score 2156.5; DB 2; Length 554;				
Best Local Similarity 73.6%; Pred. No. 2.1e-74;				
Matches 408; Conservative 67; Mismatches 78; Indels 1; Gaps 1;				
Qy	1	MNLVKAQLQKVVYKFIQDEVAIINALLEEYHNMSSESVVEYKLUKDINLTDNYL	60	
Db	1	MSVNRKQLKEMANVRFRVQDEYVAILDLAEEYHNMSENTVEYKLUKDINSLTDYI	60	
Qy	61	NYKSGRKNALKKPEYLTWEVLTKNSITPVEKHLFIWIGGOINDTAINYNQWKD	120	
Db	61	DTYKSGRKNALKKPEYLTVEILKNSITPVEKHLFIWIGGOINDTAINYNQWKD	120	
Qy	121	VNSDVTKVYFDSNAFLNTLKKITVESATNNTLESFRENLDPEFDYKFKRMETIY	180	
Db	121	VNSDYNVNVYFDSNAFLNTLKKITVESASNDLTLESFRENLDPEFNHTAFKRMPIIY	180	
Qy	181	DQKQHFIDYKSGIENPEFIIDNIIKTYLSNEYSKDLAANKYIEESLNKTTANNGDI	240	
Db	181	DQKQNFYINYKVKQKEENPDLIIDIVKTYLSNEYSKDIDELNAYIEESLNKVTENGNDV	240	
Qy	241	RNLKFADEDLVRLNQBELVERWNLAASDILRLSMLKEDGGVYLDVDILPGIQDLPFKS	300	
Db	241	RNPEEFKTEVENFYEQLVERWNLAGASDILRVAILKNTGGVYLDVDMLPGHPDLFDK	300	
Qy	301	INKPDSI-TNLSWEMIKLEAMKYKEYIPGYTSKNFDWLDEEVQSPESALSCKSKSEI	359	
Db	301	INKPDSVKTAVDMEENQLEAMKYKEYIPEYTSKHFDTLDEEVQSPESVLASKSKSEI	360	
Qy	360	FLPLDDIKYSPLEVKIAPANNVINQALISLKDVCSDLVNIOIKRYKTLNDNLNPSIN	419	
Db	361	FLPLGDIKYSPLVKIAPAKSGIINQALISAKDSYCSDDLKIQIQRKYTLNDTLGFIIS	420	
Qy	420	EGTDFNTTKKIFSDKLASISNEDNNMFMIKITNYLKVGFAPDVRSTISLGGGVYTGAYQ	479	
Db	421	QGNDFNTTWNFGESLGAJANEENISFIAKIGSVLRGVGYPEANTTITLSGPTIYAGYK	480	
Qy	480	DLLMFKDQNSTNHILLEPELRNPEFPKTKI SOLTQEBITSLWSFNQARAKSQPEYKKGIF	539	
Db	481	DLITFKEMSGIDTSLISSELNRNPEFPKTVISQATQEKNSLWQFNNEBAKQFPEYKKNYF	540	

```

QY      540 EGALGEDDNLDPAQ 553
       |||||
Db      541 EGALGEDDNLDFSQ 554

RESULT 6
Q46149 PRELIMINARY;          PRT;   2178 AA.
AC Q46149; Q46147; Q46148;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Alpha-toxin.
OS Clostridium novyi.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19402;
RX MEDLINE=95342160; PubMed=7616958;
RA Hofmann F, Herrmann A., Habermann E., von Eichel-Streiber C.;
RT "Sequencing and analysis of the gene encoding the alpha-toxin of
RT Clostridium novyi proves its homology to toxins A and B of Clostridium
RT difficile.";
RL Mol. Gen. Genet. 247:670-679(1995).
[2]
RP SEQUENCE OF 1204-2178 FROM N.A.
RC STRAIN=ATCC19402;
RA Hofmann F., Habermann E., von Eichel-Streiber C.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DDBJ databases.
DR EMBL; Z48636; CAA89565.1; -
DR EMBL; Z23280; CAA80818.1; -
DR EMBL; Z23281; CAA80819.1; -
DR PIR; S55805; S55805.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR007577; Gly transf sug.
DR Pfam; PF01473; CW binding_1; 10-
DR Pfam; PF04486; Gly_transf_sug; 1.
SQ SEQUENCE 2178 AA; 250134 MW; 9B0ADCE01C4A75A CRC64;

Query Match           24.0%; Score 2082.5; DB 2; Length 2178;
Best Local Similarity 31.0%; Pred. No. 7.4e-71;
Matches 546; Conservative 362; Mismatches 709; Indels 147; Gaps 46;

QY      3 LVNKAQLQKVVVKFRIQEEDYEYVAIINALBEYH-NMSESVVEKYKLKDINMLTDNYLN 61
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2 LITREQLMKIASIPLRKPKEPEYNLIIDALENFNRDIETGTSVKETYSKLSKLNELVDNYQT 61

QY      62 TYKSGRKNALKPKPEVLTVEVELKNSLT--VEKNLHFPIGWIGGINDTAINYNQWK 119
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62 KYPSSGRNALENFRDSLYSELRELIIKNSRTSTIAKSLSFIIWGGPISQSLEYEYNWK 121

QY      120 DVNSDYTVKPYDSNAFLINTLKKTIVESATNTLTESFRELANDPEFDYKFKRKMEII 179
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      122 MFNKDYAIRLPFYDKNSLVNTLTATIOESKVIIQONSLDGT'VGHNKFYSDRMKLI 181

QY      180 YDKOKHFIDYYKSQIEENPEFIIDNIITKLSNEYSKDLEALKNYIEESLNKITANNND 239
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      182 VRYKRELKMYENKKONS---VDDIIIFLSNVFKYDIGNKLNQKNNNNKMTAIGATD 238

QY      240 IRNLEKFADEDVLRYNQELVERWNLAASDIILRISMLEDGGVYLVDVILPGIQPLDFK 299
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      239 I-NTENILTNNKLSVYVQELIQTNNLAASDILLRIALKKYGVCCLDFLPGVNLSLFN 297

QY      300 SINPDSITNTSWEMIKEALMKCKEYIPGYTSKQFMDLBEVQRSEFESSLSSKSDKSEI 359
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      298 DISKPNGMSDNWEAAIFEAIANPEKLMNNYPYKMEQVPSEIKERILTSFVRN-HDINDL 356

QY      360 FLPLDDIKVSLEV-----KIAPANNSVINQAISLSDVSCSDIVINQIKNEYKILNDNL 414
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      357 ILPLGDIKISQELLGERLKAATGCKTFSNAFIISNNDLSLTNNLIQSLENRYEILNSII 416

```


QY 415 NP--SINEGTD--FNTTMIKIFDGLASISNEDNMFMKIMITNYLVKGFAPDVRSTINLSG 470
Db 417 QEKFKICEYDYSVSVSELVLETPKNSLMDGSSFYQOICGLSSGFKPEVNSVFFSG 476
QY 471 PGVYTAYQDLMLFKN--STNIHLEPELNFPEPKTKISOLTEQBITTSUWSNQ 524
Db 477 PNIYSATCDTYHFIKNTFMDLSSQOEIPEAS--NNLYFSK-----THDEFKSWLLRS 529
QY 525 ARAKSOPEYKGYFEGALGEDNDLPAQNTVLDKDYVSKKI--LSSMKTRNKEVHIHV 582
Db 530 NTAKEFOKLITKIYIGRTLNIEDGLNFKNKVVTSSELLKVBEVNSTKIYENYDNLML 589
QY 583 QLGDDKISYEASCNLFSPKDPYSILYQKNIEG--SETAYVYVADAEIKEDIKRYPYOIS 641
Db 590 QIQDDISYESAVNVFGKPNKSI-----IQVDDFANVYFENGIVQSDNNILNLSRN 645
QY 642 NKENIKLITFIGHGSEFNTDTFANLDVSLSEBI-----ETILNLAADISPKYIEINLL 696
Db 646 DIKKIKLITLIGHGENVFNPKLFGKTVNDLYTNIKKPLQHLEREGVILKKNYKILN 705
QY 697 GCNMFYSISABETYPGKLLIKIDRVSELM--SISQDSITVSANOYEVRIEKGREIL 755
Db 706 GCMYFPKVDINGTFVGLFNKIS---RDLQPKGFSKNQLEISANKYAIRINREGKREYL 762
QY 756 DHSGKWINKESIIKDISKEYISPNKENKIIIVKSKYLHELSTLQETRNANSSDIDL 815
Db 763 DYFGKVSNTDLIAEQISNKYVYVNEVENTLSARVEQLNKVAFKAI-----NSIOT 817
QY 816 EKKVMLTECEINVASINDROIVEGRIEEAQNLTSDSI---NVKNEFKLIESISDLYD 871
Db 818 TNNQELKQSLVNYADLITLYSELKEDIPPELDNIQIKERILNELISLHDFSNIL 877
QY 872 LKHQGLDDSHFISFEDISKTENGFRIRFINKETGNSIFIETKEIFSEYATHISKEISN 931
Db 878 FYQKNNISNMIIIFDSIIKEKYVNVKLANKITGETSVIKTYSDSLWNFTNKKYKIVD 937
QY 932 IKOTIFNVNGKLVKNLDAAEVNTLNSAFIQLSLENTTKESLNSLVAMKQVYA 991
Db 938 IKGIIVKDIINGEFIKADFEIQNPSSLNSAMLQILLIDYKPYTEILTNMTSLKQVYA 997
QY 992 QLFSTGLNTITDASKVVELSTALDEITDILPTLSEGLPIIANTIGVSIIGAIAKELSET 1051
Db 998 QIFOLSIGAQEATEIVTIIISDALNANFNILSKLVGSSVASVIIDGINLIAALTELKVV 1057
QY 1052 NDPLLROEIEBAKIMAVNLTAATAIVTSALG---IASGFSILLVPLAGISAGISLVN 1108
Db 1058 KTFPERKLEAKVGMYSIGIFLESSLISGLGATVASEILGVISVPVAGILVGLSLVN 1117
QY 1109 NELILODKATKVIDYFKHISLAETEGATLLDDKIIIMPQDVLSEIDFNNSNITIGKE 1168
Db 1118 NILVLGEKYNQILDYFSKFPYIVGKNPFS-IQDNIIPYDDIAITELNFKYKFKYGYAK 1176
QY 1169 IWRAGSGGHTLTDIDHFPSSPSITVRKPLWLSIYDVLTNKKKIDPSKDLVLPNAPR 1228
Db 1177 ISGLKVLGVTHIGENIDHYFAPSLSDYIE--LSIYPALKNDTNLP--KGNVLLPSGLNK 1234
QY 1229 VFGYEMGWTPGFRSLNDGTLKLLRIRDHV-----EGQFYWRYPAPFADALITKLKPYE 1283
Db 1235 VYKPEISAIAGANSQEGNGVELNLIRNYVDSNGNTKFPKYEAPP--EYSFYSMRVEYF 1293
QY 1284 DTNVRINDGNTRFIVPVITTEQIRKNLSYSFGSGSYSLSSIPYNNMIDNLIVENDT 1343
Db 1294 DTKVNVILDNENKTLIIPVLTIIDEMRNKISYEILGQGGQYNVILPVNQTHINIVSNKDI 1353
QY 1344 WVIVDNNVKNITTESDIOKGEIENILSKLNTIEDNKIILNHTINIFYGDINESNRFIS 1403
Db 1354 WNFVDSYIVKESKIEDNKVFLDGFINNIFSTLKVSDNGFKIGQFIS---IKNTPRAIN 1409
QY 1404 LTFSEILEDNIIILIDL--VSKSYKILLSGNCMKLIENSDD-IQQXIDHIGFNGEHQYIP 1461
Db 1410 LSFKINNNI--VIVSYTLNHEKSNSTIITSSDLNIDKNNFNLDNINIVIGLSIDNTIN 1468
QY 1462 YSIDNETKYNGFDIDYSKKEGLFTAEPFSESIIRNIYMPDGNLFIYSSKDLDIRIINK 1521

Db 1469 CIVRNDVYMEGKI-----FLNEK--KLVFIOQNELEHLYDS-----VNK 1506
QY 1522 GDVKLLIGN-----YFKDDMKVSLSGFTIBDT-----NTIKLVGVLDENGVAQI 1565
Db 1507 -DSOYLNNFINNVVKOGYIVVEGFLINSTENKYSLYIENNKIMLKGILY-ESSVFKT 1564
QY 1566 LKFMNNAKSALNT--SNSLMNLFLESINIK---NIFYNNLDPN---IFILDTN--FISGS 1616
Db 1565 IQDKIYSKEVNDVILSLIKKFTVNIQLCFPMVSGVDNENRYLEVMLSTNNKWIING- 1623
QY 1617 NSIQOFELICDKNIOPIYFINFKIKETSITLVGNRQNLIVEPSYHLDSDGNISSIVIN 1676
Db 1624 --GYME---ND-----FNNYKIVDF-----EKNVIVSGSNKLNSEGLADT-ID 1662
QY 1677 PSQKLYCIDRYNKNVIAAPLXYT 1700
Db 1663 VLCKDLENL--YIDSVIIIPKVVY 1684

RESULT 7
O68653 PRELIMINARY; PRT; 698 AA.
ID O68653;
AC O68653;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Truncated toxin A.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=CCUG 20309;
RX MEDLINE=99304497; PubMed=10376209;
RA Song K.P., Bai X.L., Chang S.Y.;
RT "Nucleotide and peptide sequences of the open reading frame encoding a truncated toxin A gene of Clostridium difficile strain CCUG 20309.";
RL DNA Seq. 10:93-96(1999).
DR EMBL; AF053400; AAC08437.1; -
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF04488; Gly_transf_sug, I
SQ SEQUENCE 698 AA; 80682 MW; 021EB268A3BDECSE CRC64;
Query Match 22.1%; Score 1916; DB 2; Length 698;
Best Local Similarity 52.9%; Pred. NO. 3.8e-65;
Matches 371; Conservative 138; Mismatches 184; Indels 8; Gaps 5;
QY 1 MNLVNAQLOKMYVYKFIQDEYVAIUNALEEHNHNSSESVVEKYLKLDINNLTDNYL 60
Db 1 MSLISKBELIKLAY-SIRPRENEVKTVLNDEYNKLTNNNNKYLQKLKLNESIDVFM 59
QY 61 NTYKSGRNKALKKFKVELTWEVLELKNLSLTPVEKNLHFVIGQINDTAINYNQWD 120
Db 60 NKYKSSNRNALSNLKDILKEVILIKNSNTPVEKNLHFVIGGVEVDIVLEYIKQWAD 119
QY 121 VNSDVTVKFYDSNAFLINTLKKTIVESATNNLTLESFRENLDNDFYKFKREMEIY 180
Db 120 INAEYNKLVYDSEAFVNTLKAIVESSTEAQLLEEIQNFQDNMKYKKEMEIY 179
QY 181 DKQKHFDIYKSGQIBENPEFIIDNIIRKTYLSNEYSKOLEALNKYTEESLNKITANGNDI 240
Db 180 DRQKRFINYKSGQINKPTVLTIDDIKSYLVSEYNRDETLLSYRTNSLRKINSNHDIV 239
QY 241 RNLEKFAEDLVRNLQBELVERWNLAASDILRISMLKEDGGVYLDVILPQIDPLFKS 300
Db 240 RANSLFTEQELNINYNQELLNRGNLAASDIVRLALKNFVGVYLDVMDLPQIHSDLPKT 299
QY 301 INKPSITNTSMEIKLEAIMKYKEVIGYTSKDNFMDLDEEVQRFESALSSKSKSEIF 360
Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYINNYTSENFDKLDQDQKQNFKLIIESKSEIF 359

Db 61 YKSSRNRLSNLKDILKEVLKNSNTSPVEKNLHFVWIGVESDIVLEIKQWADIN 120
 Qy 123 SDTVKVFDSNAFLINTLKTIVSATNTLESFRENLANDEFFDYNKPKRMILLIDYK 182
 Db 121 AYNIRLWYDSEAFVNTLKKAIVESTTEALQLEEBEQNFQDNMKFYKRMFFIYDR 180
 Qy 183 QKHFIYKSOIEENFEFIDNIITKYLSEYNSKDLKALNKYIESLNKITANNNDIRN 242
 Db 181 QKGFYKSOINKPVTITDIIKSYLSEYNRDETLESRTNSLRKINSHGIDVRA 240
 Qy 243 LEKFADEDLVRLYNQBELVERNWLAAASDILRISMLKEDGGVYLDVDILPGIDLPKSN 302
 Db 241 NSLFTQELLNINYOELLNRGNLAASDIVRLALKNFGVYLDVDMFGIHSDFKFTIS 300
 Qy 303 KPSITNTSWEMIKLEAIMKYKYPGYTSKNPDMLEDEEVQSPFSALESKSKSEIFLUP 362
 Db 301 RPSSIGLDRWEMIKLEAIMKYKYNNTSENPKLDQQLKONFKLIIESKSEKSEIFSK 360
 Qy 363 LDIKVSPLVKIAPANNVINQALISLSDYSCSDIVINQIKRYKILNDNLNPSINEGT 422
 Db 361 LENLNSDLEIKIRFALGVSINOALISKSGSYLTNIVIQKRYOFLNOHLNPAIESDN 420
 Qy 423 DFNWTMKIFSDKLASINEDNMFMUKITNYLKVGFAPDVRSTINLSGGVYTGAYQDILL 482
 Db 421 NFDYTKIFHDSLSFNSATAENSFLTKIAPYLOVGFMPPEARSTISLSGGVAYASAYDFI 480
 Qy 483 MPKDNSTNIHLPELRNFPFKTKISQTEQBITSLWFSNQARAKSQPEEYKKGVEGA 542
 Db 481 NLQENTIEKTLKASDLIEFKFENNLSQTEQINSLSWFSNQARAKSQPEEYKKGVEGA 540
 Qy 543 LGEDNDLPQNTVLDKDY-VSKILSS--MKTRNKEYIHYIVOLQGDKISVEASCNLFS 599
 Db 541 LSEDEVDNFKNTALDKYLLNNKIPSNVVEEAGSNVYHYITLQGGDDISYEATCNLFS 600
 Qy 600 KDPVSSLYKNTGSESTAYVYVAD--ABEIKIDKVRIPYQISNKRNTKLFIHGKGE 657
 Db 601 KNPKNSTIIQRNM--NESAKSHFLSDGSEILSNKYRIPERLKNKEXKVYFIHGKDE 658
 Qy 658 FNTDTPANLDVDSLSBIETILAKADISPKYIEINLL 696
 Db 659 FNTSEFARLSVDSLSNIEISSFLDTIKLIDISPKYVEVNL 697

RESULT 10
 Q9FCX5 PRELIMINARY; PRT; 553 AA.
 AC Q9FCX5; (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Toxin A (Fragment).
 GN TCDA.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1496;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C34;
 RX MEDLINE=20402122; PubMed=10931294;
 RA Braun V., Mehlig M., Moos M., Rupnik M., Kalt B., Mahony D.,
 RA von Eichel-Streiber C.
 RT "A chimeric ribozyme in Clostridium difficile combines features of
 RT group I introns and insertion elements."
 RL Mol. Microbiol. 36:1447-1459(2000).
 DR EMBL; AJ131844; CAC03681.1; -.
 DR InterPro; IPR007577; Gly transf. sug.
 DR Pfam; PF04488; Gly transf. sug; 1.
 FT NON-TER 553
 SQ SEQUENCE 553 AA; 63913 MW; DDE4551A6D8C3B25 CRC64;
 Query Match 17.0%; Score 1474; DB 2; Length 553;
 Best Local Similarity 51.7%; Pred. No. 1.7e-48;

Matches 287; Conservative 108; Mismatches 158; Indels 2; Gaps 2;
 Qy 1 MNLVNAQLOKVVYKEROEDDEVAILNALAEYHNMSSESVVEKYLUKLDINLNTDYL 60
 Db 1 MSSIKKELIKLAY-SYRPRENEYKTLTLDDEYNKNTNNENKYLOLKLESIDVFM 59
 Qy 61 NTKYKSGRNKALKKFKLEYLTWELKNNLUTPVEKNLHFVWIGQINDTAINYNQMD 120
 Db 60 NKYKSSRNRLSNLKDILKEVLKNSNTSPVEKNLHFVWIGVESDIALEVIKQWAD 119
 Qy 121 VNSDYTVKVFDSNAFLINTLKTIVSATNTLESFRENLANDEFFDYNKPKRMILLIY 180
 Db 120 INAEYNKLVYDSEAFVNTLKKAIVESTTEALQLEEBEQNFQDNMKFYKRMFFIY 178
 Qy 181 DKQHFIDYKSOIEENFEFIDNIITKYLSEYNSKDLKALNKYIESLNKITANNNDI 240
 Db 179 DRQRFYNYKSOINKPVTITDIIKSYLSEYNRDETLESRTNSLRKINSHGIDI 238
 Qy 241 RNLEKFADEDLVRLYNQBELVERNWLAAASDILRISMLKEDGGVYLDVDILPGIDLPKS 300
 Db 239 RANSLFTEQELLNINYOELLNRGNLAASDIVRLALKNFGVYLDVDMFGIHSDFKT 298
 Qy 301 INKPDSTNTSWEMIKLEAIMKYKYPGYTSKNPDMLEDEEVQSPFSALESKSKSEIF 360
 Db 299 IPRSSIGLDRWEMIKLEAIMKYKYNNTSENPKLDQQLKONFKLIIESKSEKSEIF 358
 Qy 361 LPLDDIKVSPLEVKIAPANNVINQALISLSDYSCSDIVINQIKRYKILNDNLNPSINE 420
 Db 359 SKLENLNSDLEIKIRFALGVSINOALISKSGSYLTNIVIQKRYOFLNOHLNPAIES 418
 Qy 421 GTDFTNTMKIFSDKLASINEDNMFMUKITNYLKVGFAPDVRSTINLSGGVYTGAYOD 480
 Db 419 DNNFTDTTKIFHDSLSFNSATAENSFLTKIAPYLOVGFMPPEARSTISLSGGVAYASAYD 478
 Qy 481 LLMFKNSTNIHLPELRNFPFKTKISQTEQBITSLWFSNQARAKSQPEEYKKGVE 540
 Db 479 FINQENTIEKTLKASDLIEFKFENNLSQTEQINSLSWFSNQARAKSQPEEYKKGVE 538
 Qy 541 GALGEDNDLPQNT 555
 Db 539 GSPSGDNGVDNFNKT 553
 RESULT 11
 Q9ZGR4 PRELIMINARY; PRT; 3169 AA.
 AC Q9ZGR4;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative cytotoxin.
 GN L7095.
 OS Escherichia coli O157:H7.
 OG Plasmid pO157.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EDL933;
 RX MEDLINE=98391744; PubMed=9722640;
 RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
 RA Blattner F.R.
 RT "The complete DNA sequence and analysis of the large virulence plasmid
 RT of Escherichia coli O157:H7."
 RL Nucleic Acids Res. 26:4196-4204(1998).
 DR EMBL; AF074613; AAC70163.1; -.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR InterPro; IPR007577; Gly transf. sug.
 DR InterPro; IPR006473; Peptidase_C38_yt.
 DR Pfam; PF04488; Gly transf. sug; 1.
 DR TIGRFAMs; TIGR01586; yopT_cys_prot; 1.
 KW Plasmid.

SQ	SEQUENCE	3169 AA; 361971 MW; C2CDB48A72BD5FD1 CRC64;	
	Query Match	5.9%; Score 512.5; DB 2; Length 3169;	
	Best Local Similarity	20.2%; Pred. No. 4.6e-11;	
	Matches	422; Conservative 333; Mismatches 720; Indels 619; Gaps 104;	
QY	14	YVKERI--QEDVAVAILNAL--EYHNMSSESVVYKYLKLDINLNTDNLNLYTKKSGR 68	
DB	160	YIKIRKTRGABDQTTITQSLIINELLNGVDRNTI-PFOKISELNDIHSYENQJKNR 218	
QY	69	N--KALKKFEYLWVLELKNS--LTPVERNLFHFWIGGQ 106	
DB	219	KGIELYKQGLLSSLNDKNGKQSLDNASKIINILGIEVQSHKVDIEFFIHAVVAGA 278	
QY	107	INDTAINYINQWQNSDYTVKPYDSNAP--LINTLKT-- 144	
DB	279	PPDNTFSYITAFNTYKDYTYLLWIDPNAFAAKFSGILKNIAMYAIMELRRTNPHLAE 338	
QY	145	---IVESATNTLE--SPRENANDPEFY--NKFYKRMELIYDKQKHFI 167	
DB	339	EMNEVILKIONIETIEFKETRELRKLELNRYSKLSLTKETKEPNVFFLESMTGMQDNFY 398	
QY	188	DYKSOIBENPEFT--IDNLIK--TYSNEYSKDEALNKYIEBSLNKIITANNGN-- 238	
DB	399	TYCISNGISNTDDSRDLFLNVLKLSPEVQNDPKSTVEKNKRDIDLKNTISQKGDPR 458	
QY	239	---DIRNLEKFADELVELYNQELVERWNLAASDIILRISMLKEDGGVYLDVILPGIQP 295	
DB	459	QLRDINTLESFKKQDFFYQOEMLRWNYAAASDQVRINILKEYGGIYTDILPAYSD 518	
QY	296	DLFSKINKPDSITWSEMIKLEAIMKY--EYTPG--YTSKNPMDLDEEVQSFESALS 351	
DB	519	KVSQIINE-KSDKRRFFEDLRLRIISGISILSKGKYSIKH-DGLDETTLAQLNAIL- 575	
QY	352	SKSDKSEFLDLDIKVSPLEVKIAFANNVI--NQALISLKDSYC 395	
DB	576	SEIEK---LTIDY-FKPEVETKVVRDTFKFYKQKWENTWNRGNMFMLTHKSKC 630	
QY	396	SDLVINQIKRYKIILNDNLNINSGTDFNTMKIFSKLASIGNE-----DNM 444	
DB	631	IDFILSGQKQVLEL-QRIRNISIYNNLFYTT-----EDLSLNVAIGGIPAKKYLEHG 684	
QY	445	MFMIKITNLYKVGPADVRSTINSGP----- 471	
DB	685	LF---SEYRQDGTIPYVSTINISGPDIMQMKYKSLGRIGEVHAIKONKLSVDNFL 740	
QY	472	GYTGAYQDLMF---KNSNINHLEPE-----LNFEPPTKISQLT-----EQBIS 518	
DB	741	GYASSNKDNKSFNWLNPVSVGINDITPDDESSWAVRNNDINKILFEKINCHVEKLPTS 800	
QY	519	LWSFNQARAKSQFEYKKGYPEGALGEDDNLDFACNTVLDKDYVSKILSSMKTNRKEYI 578	
DB	801	LY-----YEIDSRFFQGW---DNKSIKHVTEINKDLI--KQINLLTSSNIDV 844	
QY	579	HYIVLOQDKISYEASCNLFSDPYSSILYQKNIEGSETAYYYVADAEIKEDIKRIPY 638	
DB	845	KLLIKL--DRELYAISSKI-----DNPALARSIRTLQLQANVTSNTPEPNTINFIY 896	
QY	639	QISNKN-----IKLTTIGHKSEFNTDTFANLDDVS-----LSSEIET 677	
DB	897	DFYRKQDLLSAIKL-----FSRNDADTKIIWVNSVMKRVFLREVISC 942	
QY	678	ILNLAKADISPKYIILNLCNMPYSISAEETYPGKLLIKIKORVSELMPSISQDSI-- 735	
DB	943	VLRKKVD--SYINEN-----KXNLSKEDA--GALRDYAKLQKHLEFNLDDDDGYKK 990	
QY	736	TVSANQYEVRIINEEGREILDHSGKWINKESIIKDISKEYISNP-----K 783	
DB	991	IITWAY--IKERKL-----SGIYNIENSIISGHESFDIIRSNQHEWGLDSTVBQPK 1042	
QY	784	ENKIIVSKYLHELSTLLQEIIRNANSDDIDLEKKVMLTECHINVAS-----NIDRQIVE 838	
DB	1043	KFEFYKSE-LSSAKSIPDDIKNK-YITDPETKRNVLHQDSDIKERIAFLIDISHVAYP 1100	

QY	839	GRIEAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHQNL 878	
DB	1101	GLLEKLOLS---GYFSDINIIEAYLLASGYGSHYGVVYAPSDKLELLRRHTK 1156	
QY	879	DDSHFISPEDISKTENGFRIRINKTCNSI---FIETEKEIFSEYATHIKSINIKDT 935	
DB	1157	SNSEWI--EKITP---VYDILSDNSVNLRPPLSEBQKILNDIKLEISKSVS--EQ 1207	
QY	936	IPDNVNGKLVKNVNDAAHEVNTLNSAPFIOSLIYNTTK-----ESLSNLSV----- 983	
DB	1208	YFMKLTQEKSSVIGIKYSVDFRYNENFLSLPQNQLTLPFMYRYFEMLYDIHIGIEN 1267	
QY	984	AMKVQVTAQLFSTGLNTITDASKVVELV-----STAL-DEITDL 1021	
DB	1268	KANREFTYKSKESLMDFLINDERVNLGLIKKYKYLSEIHRILTNTNSFPADISPL 1327	
QY	1022	LPTLSEGLPIIATII-----DGVSLGAIAK-----ELSEINDPLLR 1057	
DB	1328	LOTI---CPSITIIKKTEYVGHQLTNMTAVSVKPYDPFNSLGAINSIDKSDVPAH 1384	
QY	1058	QBIK-AKIGIWNVJTAATAIVTSAL-----GASGFSILL 1093	
DB	1385	TIVEQAKYNLLSWNDFYNTHASIWDTIARQHKSTNIEPHQSLDFDRDSKGLGLSLY 1444	
QY	1094	VPLAGISAGISLVNN---ELLQDKATKVIDYFKHISLAETEGAFLLDDKIMPODD 1149	
DB	1445	LDTGYYGGYQKLRNIDTASTLYOTK-----YNDNLKSNRODFFLRKTQRIITWNE 1498	
QY	1150	LVLSEIDFNNSITLGKCEIWRAGSGHTLDDI--DHFFSSPSITYRKPLWLSYDVLN 1207	
DB	1499	L-----GNRLKNAQLEVELEK---DPILTEGLIYQRRISSLLITTEYHSLAQQLSS 1548	
QY	1208	IKK-----EKIDFSKDLMLNAPNRFVGYEMGWTEGFRSLDNDGTKLLD 1252	
DB	1549	FRVTDNPFHGHCDPHSLAQLTFIKNI-----TSNR-----NFSLSYSGSI----- 1589	
QY	1253	RIRDHY-EQGFVWRVAFIADALITKLPREDIVNRINLDG---NTRSFIVPVITTEQI 1308	
DB	1590	-KVIFSESLNNWKYIKLPLVQTSLLRDYLTTPKLSSTGGSNINMGHLVPV----- 1642	
QY	1309	RKNLSVSP-YGSGGSVLSLSLSPYNNIDNLNVENTWIDVNVKNTTIESDEIQ----- 1363	
DB	1643	-----SFIYDIGV-----INGRISST---DVNKIRSRKINGDILQHOVN 1682	
QY	1364	-----KGELIENILSKLINEDNKI-----ILN---NHTINFY- 1392	
DB	1683	THYLSSECTQKIDVIDFLIGIODNTIKVLESIDIKPISEIQPLHSILSRQEHVKNLLS 1742	
QY	1393	GDINE-SNRP---ISLTFSEILEDINI-----IIBI---DIVSKSYKILLSGNCMKL 1436	
DB	1743	GULDEFSNKLKQGLSLKTNVLSVNNFKESKINSITVEVTVDLQGLRYVDIDTRVIGL 1802	
QY	1437	I--ENSSDLOQKIDHGFNGEHQKVPYSYIDNTEKNGFIDYSK---KEGLFTAEPN 1490	
DB	1803	TFKEGINSISEALEHMN-----IDAIMSVIGLVQYARMIKNNONISAIIDAG 1849	
QY	1491	B-SIIRNIWYPSNNLFTYSSKDLIRINKGVYK-LLIGNYFKDDMKVLSLFTIEDTN 1548	
DB	1850	AVSDIKNV-----DKFLGGILTLTNRYNPGVSGASLEGFISGSEVCAS----- 1897	
QY	1549	TIKLANGV---YLDENGVAQILKFMNNAKSALMTNSLNFLESINIKNIFYNNLD----- 1600	
DB	1898	--RMGGTAGRYL--SNVAKVIL-----PLLDI--GINWSLYDSSLSNEAKAT 1939	
QY	1601	PNTEFI---LDNPF-IISGNSIGQFE-----LICDKNKIQOYFINKFKETSILYV-- 1650	
DB	1940	TQIEYISTAIDVSFSSINTALSIGAAYPPLAIA-----IVBITI-FSEVKNYAVYVNO 1993	
QY	1651	-GNRONLIVEPSVHDD-----SGNISSTVINFSQKLYG---IDRYNKKVII 1694	
DB	1994	INERHKLWEAEKYLDNGSAKVLINKATGIIDLNNQVLGNLYDMRENPPIL 2047	

```

RESULT 12
082916 PRELIMINARY; PRT; 3169 AA.
AC 082916;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Toxin B.
GN TOXB.
OS Escherichia coli O157:H7.
OG Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7;
RX MEDLINE=99290540; PubMed=9628576;
RA Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,
RA Yutsudo H.C., Kubota Y., Yamaichi Y., Iida T., Yamamoto K., Honda T.,
RA Han C.G., Ohtsubo E., Kasamatsu M., Hayashi T., Kuhara S.,
RA Shinagawa H.;
RT "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an
RT enterohemorrhagic Escherichia coli O157:H7 derived from Sakai
RT outbreak.";
RL DNA Res. 5;1-9(1998).
DR EMBL; AB011549; BAA31815.1; .
DR PR; T00296; T00296.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR007577; Gly transf. sug.
DR InterPro; IPR006473; Peptidase C58_yt.
DR Pfam; PF04488; Gly transf. sug; 1.
DR TIGRfams; TIGR01586; yopT_cys_prot; 1.
KW Plasmid.
SQ SEQUENCE 3169 AA; 361943 MW; D9BED726A62DDBFF CRC64;

Query Match 5.9%; Score 507.5; DB 2; Length 3169;
Best Local Similarity 20.08; Pred. No. 7.1e-11;
Matches 420; Conservative 338; Mismatches 716; Indels 621; Gaps 105;

QY 14 YKFRP---QEDFYVAILNAL--BEVHNMSESSVWEKYLKLDINNLNLYNTYKSGR 68
DB 160 YIKRTRGAEQDTTITQSLIINELLANGVDRNTI-PFQKISELNDIHSYENMQIKNGR 218
QY 69 -----NKALKKFXE-----YLTWEVLEKNSLTPVEKNLHPFWIGG 105
DB 219 KGIELVKQGLSSLLINVKNGKQLSDNASKIINLLGIEYQSHKV-DIEPFIHAVVWAG 277
QY 106 QINDTAINYNQMKDVNSDYTKVFXDSNAF-----LINTLAKT----- 144
DB 278 APPDNTFSYITAFUNTKYKTYLLWIDPNAFGAAKPSGILKNIAINYAIMRLRRTNPHLA 337
QY 145 -----IVESATNNTLE--SPRENLDPEFDY-----NKFYKRMELIYDKQKH 186
DB 338 EEMNEVILKQNIQNEIEKEFERLEKLENYKSLTSEKENVFPFLESIMGMQDNY 397
QY 187 IDYYSQIENPEFI-----IDNIK--TYLSNEYSKDLBALNKYIEESLUNKITANNGN- 238
DB 398 FTYCISNGISNTDDISRLDPLTNVLSPEVQNDFKSTVEKNKRDIDLLKNTISQKFGDR 457
QY 239 -----DIRNLEKFADEDLVRNLQELVERNLAAASDILRISMLKEDGGVYLDVLDLPGIO 294
DB 458 FQLRDLINTLESFKKQPQDYFFQEQEMLLRWNYAASDQVRINILKEYGGIYTDILPAYS 517
QY 295 PDLFSKINKPDSITNTSWEIMKLEAMKYK-EVIPG--YTSKNFMDLDEVOQRSFESAL 350
DB 518 DKVQCIINE-KSDDKRFEDLKRLRIISELSILKGEKYSIKH-DGLDRTTNQLNML 575
QY 351 SSKSKDSBIFLPLDDIKVSPLEVKIAFANNSVI-----NQALISLKDSY 394
DB 576 -SEIEK---LTIDDY-FKPVETKVRDTPKIPRYQKWNTWNIRGNNNFMLTHKGSK 629

```

```
Db 1590 ---VKIYFSESLNWKYIKLPLVQIGSLRLDIYLTTPKLSFGSGSLNIMGLHPV----- 1642
Qy 1308 IRKNSYSF-YSGSGSYSLSPNNMIDNLVENDTWIDVNVNKNYITSEDIQ--- 1363
Db 1643 ---SFIYDIGV-----INGNRISST---DVQNKIRSLKINGDILQHYI 1681
Qy 1364 ---KGBLIENILSKLNIENKI-----ILN---NHTINFY 1392
Db 1682 NTHYLSSECTQKIDIVDFLGIQDNTIKVKLESIDKPSBIQOPLHSLRQKEHVKNLL 1741
Qy 1393 ---GDINE-SNRP-----ISLTFSTLEINI-----IIEI---DLVSKSYKILSGNCKM 1435
Db 1742 SLLDEPNKLRKQGLSLKTNVLSVNNFKESKINSDTVEVTVDLQGLYFVDDITRIG 1801
Qy 1436 LI--ENSSDIQKIDHIGFNGEHQKYPISYVIDNETKNGFIDYSK-----KEGFTAEFS 1489
Db 1802 LTFKEGINSLSALEHMN-----IDAIMSVIGLVQVARMIKMNDNISADHA 1848
Qy 1490 NB-SIIRNIMYDPSNNLIYTSKDKJRIINKGDYK-LLIGNYFKDDMKVLSFTIEDT 1547
Db 1849 GAVSDIKNIV-----DKFLGGILTLTNNRVNPGVSGASLEGFTSGLEVCA5----- 1897
Qy 1548 NTKLNGV---YLDENGVAQILKFNNNAKSALNTSNLWNFLSINIKNIFNNLD--- 1600
Db 1898 ---RMGGTAGRYL--SNVAKVTKL-----PLLDI--GINIWSLYDSSLNHAKA 1938
Qy 1601 -FNIEFI---LDTNF-IISGNSIQGF-----LICDKNIQPYPINFKIKETSYTLV- 1650
Db 1939 TTQIEYISTAIDVSSINTALSIGAIYPLAIA-----IVPITI-PSHEVKNVAVYVN 1992
Qy 1651 ---GNQNLIVPSYHLD-----SGNISVTINFSQKLYG-----IDRYNVKVII 1694
Db 1993 QINERHKWLEAEKYLDNGSAKVL5INKATGIIDLSNNQVLNIGIYLDRENPPIL 2047
```

RESULT 13

```
Q8EWP8
ID Q8EWP8 PRELIMINARY; PRT; 3317 AA.
AC Q8EWP8;
DT 01-MAR-2003 (rEMBLrel. 23, Created)
DT 01-MAR-2003 (rEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (rEMBLrel. 25, Last annotation update)
DE Predicted cytoskeletal protein.
GN MYPEL550.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Saiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
RW ENBL; AP004170; BAC43946.1; -.
KW Complete proteome.
SQ SEQUENCE 3317 AA; 385840 MW; 0641151BAFA992BB CRC64;
```

Query Match 5.7%; Score 494; DB 16; Length 3317;
Best Local Similarity 20.1%; Pred. No. 2.5e-10;
Matches 419; Conservative 355; Mismatches 701; Indels 610; Gaps 105;

```
Qy 14 YKFRIQ--EDEVYVAINALEEYHN-MSESSVVEKYL-----KLKD-----INNLTD 57
Db 583 YINVLKLERQFSKIANMLEPIANKLSGDSVNEENFGQFEGKINKIDDLISKIASERE 642
Qy 58 NYLNTY-----KSGRNKALK--KFKEY 78
Db 643 NYENTYSSVILEVINANEAHKENVLNLIKTEBKNNLISESKSAQNSKLGELEL 702
```

```
Qy 79 LTME-----VLEKXNSLTPEKNL-----HPWIGGOINDPAINYNQWKD--- 120
Db 703 IGLQNEETESLDEKXQFISDVEQLLIEKREKLNDEKILDKNSINDVMLNFKNEKINDLE 762
Qy 121 -----VNSDYTVKVPYDSNAFLINTLTKT-----IVESATNWTLESFRENLDNPE 165
Db 763 YGFDRLKATINLSDSLATNENQKELFNKIBETLVNNYQIVKEEVKQDLTOILNNVNLN 822
Qy 166 FYNKPYKRMELIYDKQKHFDYKSOIEENPEFIIDNIITKLSNYSYKDLKALNKYI 225
Db 823 EAYGSLNQOGW--TYDKGFNDID---SQFEQ-----FRYLLKDVTENL--INKNT 865
Qy 226 E--ESLKNITANNNDIRNLEKFADEDLVRLYNQOELVERWNLAASDILRISMLKEDGV 283
Db 866 EYDENANSVNNISDNFXDLISGIKEENVK-YSEDIK--NVAETNLRNKGIEFGEBI 922
Qy 284 YLDVILPQIGPLPKFSINKPSITNTSWEMTKLEAIWKYE---YIPGYSKFNMDLDE 340
Db 923 ---SEKISNLVADIEKNKVAESNQANTINYILETNLANQKELVGLIKNIESNDKFLKE 979
Qy 341 EVORSEFESALSSKSKSEIFPLDDIKVSPLEVKIAPANNVINOALISLKDSYCSDLVI 400
Db 980 VDKRHR-----LDNLKUNELDELISLQNSEI--QALIEKNDFIHDLEV 1022
Qy 401 NOIKRKYKILNDNLNPSINEGTFDNTTKIPSDKLASIS-NEDNMFMKI----- 450
Db 1023 FILEKQNFNDENINELKQSAQEI---TELLNKNKEINLENFENRLREIDHSFKSELG 1079
Qy 451 -----TNY--LRVGFAPDVRSTIN-----LSG 470
Db 1080 NQNDPFRNLEQQLTSNYEIIKNELNNDIVQVINNVNELSSVSVDRIENLSFKDFSKPDS 1139
Qy 471 PGVYTGAYQDMLLFKDNSTNIHLPELANPFPFKTKISQLT---EOETISLWSFNQARA 527
Db 1140 IDIQDPSFRHL--KDVTE--LIDKNSNYENINNVNMDLNTKFBSTIKLSEKNRTV 1195
Qy 528 KSOFEVYK-----GYFEGALGE--DONLDPQONTVLDKDVSKKILS 568
Db 1196 RKLNKEIKKTNWTKNKIDINSIEIGDRFANLIDEIKQKNKF-ENSLDQVY---SIIE 1251
Qy 569 SMKTRNKYIHY-----IVQLQDKISYZEASCNL-----FSKDPYSSIL 607
Db 1252 ENKSHOREILNYIKNLEENNEFISQIEAQKDNE--NLKLEBELISLQNEETESLL 1308
Qy 608 YQKNIEGSETAYVYVADAEIKEDKIRIPYQISNKNIKL-TFIGHGKSEFN----- 559
Db 1309 YEKV-----DFINDIQAALLNKN--KLDNENIIRLEDSINNALGNFNDKISELE 1356
Qy 660 ---TDTFANLVDVSLSEIETILNLAKADISPKYIEI-----NLLGCNFSYSISAEFY 711
Db 1357 QNFNNKANLIDLSLSEID-----GKQVEFFDSVQNSLSHN---YELVKNE-- 1400
Qy 712 PGKLLLIKIDRVSELMPSISQDSITV-----SANQYEVRIINEEGKREILD---HS 758
Db 1401 -----FKDLSLSILSNLNDVNTVENTNSKIGDISSEIRDISLNNOQAKFSDITQFDS 1453
Qy 759 GKWINKE--ESII-KDISKEYIS-----FNPKNKIIYKSKYLHELSTLLQEIERNANS 810
Db 1454 FRHLKEVTENLIDKXSSNVENINIMDDINTFESLITKLR--NENNRVTKLNKEIKK 1511
Qy 811 SIDLKXKVMLTECEI-----NVASNDR---QIVEGRIEEAKNLTSDSINTYI 855
Db 1512 TNLRTNKKLNTTISEIGDKPFDLIQIKQKNKNFESSQLDVIYSIIEENKSHQREILNYI 1571
Qy 856 K-----NEF-----KLJ-----ESISDSLYD-----LK 873
Db 1572 KNLDENNNEFINQIEAKYNNENKLEBELISLQNEETESLLYKKNDFINEVETFLN 1631
Qy 874 HONGLDSDHSFISPEDISKTEGFRIRFINKETGNSIFIEETEKEIFSEYATHI-----SKE 928
Db 1632 KKKLDNENINRLED---SVNNAIVFNKK-----ISALEENFNSRLANIDLSSISEE 1681
Qy 929 ISNIKDTIFDNV-----NGKLVKKNVLDAAHEV-NTLNSAFFIQSLIEYNTTKESLNL 981
```

Db 1682 IDGQVFPDVSQSSLSYVVELVKNFQKDSLNILNIN-----EVNTVENVTNK 1732
Qy 982 SVAMKQV-----YAQL-----FSTGINTITDASKVELVSTALDETIDLLPTL 1025
Db 1733 IADISTEIKDINLNQSRFSDSOPNFTLLKEVTEG--FINKENSHYDELANNITSI 1790
Qy 1026 SGLP-IATIIDGV-SGA-AIKELSENDDLLQCEIAKIGIMAVNLTAATAIVTSA 1082
Db 1791 DESFKLISFQKSTNQIGENLVKDKHNSKHL---AIKNDSSIMEN-----1835
Qy 1083 LGIASGFSILL-----VPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETGPTL 1138
Db 1836 -----SKFDLIEBMKQTKDIENSVDTLN--WLIESNKVQOEIFNVVSLLE-KSTSSL 1888
Qy 1139 LDD--KIIMPQDVLSE-----IDFNNSITIGKEIMWAEGSGHTLTDIDHFPSSPS 1192
Db 1889 VNEMQAQLDKKIKLKELETLIDLQNSEI-----ESLIDDKNDPIE---1930
Qy 1193 IYTRKFWLSIYDVNLTKKIDPSKDLV-LPNA-----PNRVFGVEMGTGPFSL 1243
Db 1931 -----EIEFNQKKEI-FNKENFVKLENSIGNISAFDXXKITDLFNFDRKLSNL 1980
Qy 1244 DNDGTLLDRIRHYEGQFWRYFAFIADALITKLPYEDTNVRLNDGNTSRPIVPI 1303
Db 1981 DLSLKNLFDKDD-----STLASVINSVEDNLNSNVEILNELKELN-----2023
Qy 1304 TTEIRKINSYSFYSGGYSLSLSPYNNIDNLVENDTWID-VDNVYKNITTESDPI 1362
Db 2024 ---KILEDI-----SEANSINEINEQPHKLEEVNSQFQDQFVLTWT--EDII 2069
Qy 1363 QKGE-----LIENILSKLNTEDNKILNHTTINPYGDINSNRPISLTFSLIEDINIIIEI 1418
Db 2070 HKSENQIVIDDLLEKINDEFRLIQNLRSNELNSLMDKFNELNSGLVERFN--SNF 2127
Qy 1419 DLYSKYKILLSGNCKLIENSDDIOQ-KIDHIGF---NGEHOKYIPYSYIDN-ETKN 1472
Db 2128 DNINNKFEFIA---KLDEKTSFENRMDLNFVYELNKSYQKNIS-SYIQKIESNN 2182
Qy 1473 GFI-----DYSKGEGLTAFSNEISIRNYMPDSNN-----LFIYSKOLK 1514
Db 2183 EPIISDLDARTKENDKLEKLEVLIDIQNEIVRLI---DENNKYIDEVEYFLTKKEAL 2239
Qy 1515 DIRIIN--KGDVKLLIGNYKDDMKVSLSTIBDTNTIKLN--GVYLDE-----1559
Db 2240 DNKKLNDIESSINNINNNFEKIKLGENFN-EKVDSIELQNLGNLLESDKTKLDVIRS 2298
Qy 1560 -----NGVAQILKFWNNAKSAL-----NTSNLSMNTLESINK-NIFYNN 1598
Db 2299 QLSKNVYELLSEEFKNFASLIKSIENLYSSYGKDSLEQIASLKFCEYINEGISLIGS 2358
Qy 1599 LDPNIEFI---LDTNFIISGNSIGOFELICDKD---KNIQPYFINFKIKETSITLYVG 1651
Db 2359 INKNFDFLIKELETKYKILASQVDMNLILOSNKINSKEIVSYENLGNVANUSQKEK 2418
Qy 1652 NRQNLIVPSYHLDDSGNISSTVI-NFSQKLYGID---RYNKNV 1692
Db 2419 EKQALIFKIFDELNEINIIONQIGSLLEEDYIIDEVQKFTNKI 2463

RESULT 14

Q814R2 PRELIMINARY; PRT: 2792 AA.
ID Q814R2
AC Q814R2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Reticulocyte-binding protein, putative.
GN FL2520W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyse S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian L.M., Mungall C.,
RA Venter J.C., Carucci B.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AB014852; AA036586.1; --
SQ SEQUENCE 2792 AA; 328943 MW; 2952C3F104237112 CRC64;

Query Match 5.6%; Score 489; DB 5; Length 2792;
Best Local Similarity 20.1%; Pred. No. 3.1e-10;
Matches 407; Conservative 322; Mismatches 685; Indels 614; Gaps 96;

Qy 1 MNLVNAKQLOQM-----VVPFRIOEDEVAI-----LNAL 31
Db 856 LSTVNS'ELNKLNRNFFSEKDDLNVESKVKQDENIFIEKNKIFDDIHKIDELYNKKTNAI 915
Qy 32 EYHN-----MSSESVVEKYLKLD--INNL-----TDNLYNTYK---KSGRNAL 72
Db 916 KNLNALINGSNNLSLSDSVMKNGKDIINELRSORYLIQDNFDIYKEKIFLKNLKNGL 975
Qy 73 KKFXYLTMEVL-----ELKNSLTPVEKNL-----HFIW--IGGOIND 109
Db 976 EEENRLSNTYMKELKTEAKQEKYKLEKNINTYDDTFLEKLGDNVEWELKIELNG 1035
Qy 110 TAINY-----INQMKDVNSDYTVKVFVDSNAFLINTLKTIVESATNNTLESFR 158
Db 1036 LNVNYNLQANIDTLILKPYID-HIDHISLIESLKHNIENKIKVPLNERLAKDFIQTK 1094
Qy 159 ENLNDPFDYKFKRMEIYYQKQPIDY-YKSOEENPEFIDNIITKLSNVEYSKD 217
Db 1095 FNTNDIKLDHNLITIRI---DNR---DYNHMKLEEKEDLFKNI-----NDKKEE 1140
Qy 218 LEALNKYTESLNK-----ITANGNDI-----RNLEKFADE 249
Db 1141 IEKLLKLEENKNNKMENTNISYGIKKKNLIDIESMSLSLKSITTDENLYKLQNV 1200
Qy 250 DLVR-LYNQELVERWNLAAASDILRISMLKEDGGVYLDVILP---GIQPDLPKSIKNPD 305
Db 1201 DEFRLLVILIEINEKIKNOIKEASVERVEIELYKEKILSMNSIKDISK-----1254
Qy 306 SITNTSWEIMKLEAIMKYKEYIPGYTSKNFMDLDEEVQVSFESALSKSKSEIFLDD 365
Db 1255 -LTTFKYNEIYDECIITKENNIKELYT-KSSSLLEE-----SCRDKN-----MDI 1296
Qy 366 IKV--SPLE--VKAFANNSVINQALISLSDYSYCSDLVINOIK-----NRYKILN 411
Db 1297 IKTKSVLDYELTKTSIQNNNDIKNSLASLKMYA---ILOSIKLDVNTKYILDRSYKC--1351
Qy 412 DNLAPSINEGTDFNTWKIPFSDKLASINEDNMFMKITNYLKVGPADVRSTINLGGP 471
Db 1352 -----EDHARDLEMELEKESALSKN-----IK-----LKIEKAEFTRKV-----1386
Qy 472 GVVYTGAYQDILLMFKDNSTNIHLLEPELNFNFFPKTKISQLTEQBITSLWSF--NQARAKS 529
Db 1387 -----LGSEKHESIDY-----IFNIEKIQDISVIESDITMCIONAYDNKTKSML 1432
Qy 530 QFEYKGY-PEGALGEDNDLDFQNTVLDKDYVSKKILSSMKTRNKYHYHYVQLQGDK 588
Db 1433 HFQNVHRGIDLIGILNKKN-----QGVISKPAESGNI-----SEBY-----EQGN 1472
Qy 589 IS-YEASCNLFSK---DPYSSILYQKNTEGSETAYVYVADAEIKEDK-----633
Db 1473 INDYQKCKTYSEASDNYDEISVRKD-----SLEFEKKITNILDVLI 1517

QY 634 YRIPYOISKRN-----IKLT-FIGHKGFEFNTDTFANLDVDSLSSEIETILNLAKA 684
 Db 1518 FNMKTELEKOSVNNILGEMKLTASINEKSHVLSKIRVLRNQNKIKDDELLNNEKS 1577
 QY 685 DISPKYIEINLLGCNMFYSISABETYPGKLLKID--RVSE-----LMPSI 730
 Db 1578 KIAVEQFEL-----YMGKLEVAIPDINRLHDKAERMFSAQFLMKPML 1620
 QY 731 SQSITVSANQYVRINEBGRILDRHSGKWNKEESIIKDISKEYISFNPE-----N 785
 Db 1621 DMSLILELKNLEBK-----NKEHDYENI---EYIQDEEKHKYELT 1660
 QY 786 KIIIVSKYLHELSTLLOEIRNANSDDIDLEKKVMLTECEINVASNIDQIVEGEIEAK 845
 Db 1661 KLIEVIOFVDNIENLEKRYKYEQGNLE-----NVYNSNN--IKDRIEETK 1706
 QY 846 NLTSDSINYIKNEFKLTESISLYDLKHONGLDLSHFISFEDISKTEFGFIRFINKET 905
 Db 1707 NLSVLVN-----FPSSIKNIIYLKHS-----VKMEDNS----- 1737
 QY 906 GNSIFITEKEIEFSEYATHISKEISNIKDTIFONVNGKLVKVNLDAAHEVNTLSAPFI 965
 Db 1738 ----YIKKMDIYDEF---MESYNLIQKTIIESSN---DDIEYBELQVNRNIEKFEI 1785
 QY 966 QSLIEYNTTKESLNSLVAMKVQVYIAQLFSTGLNTITDASKVVELYSTALDETIDLPTL 1025
 Db 1786 QLV-----KABEDMKLYWNSIKNDVYKLVYIKDMLLELNRCKSVDTILNEGDFINCK 1841
 QY 1026 SEGPIIATIDGV--SIGAAI---KELSETNPDLRQETAKIGMAVNLTAATAIYV 1080
 Db 1842 AKENSVSGD--DDVYNELNKAINKYKEIHEKSNFVCKNEAESLFGIIVKSSNIIGKIIIT 1900
 QY 1081 SALGIASFISILLVPLAGISAGIPSLVNNEL-----ILQDKATKVIVDFKHI 1127
 Db 1901 -GLG-----LEKEDVDLGTSLLNSSLHFHTASINKLYSTIESDVYDKVNDCLKS- 1950
 QY 1128 SLATEGAFTLLDKIMPODDLVLSEIDFNNSNITLCKEIKWRAEGSGHTLTDIDHF 1187
 Db 1951 SLDIVKYSFDIEKKK---RQINSIMEEINNVAHITI-KIE-----LANMINDS 1995
 QY 1188 PSSPSITVRKWLSDIYVNLNKKKIDPSKDLVLPNAP--NRVFGYEMGW----- 1236
 Db 1996 RNKISVNLKIVYSINKIKNVKWTCDSSVHMIKEGVEDKLUKEYKYNVEEKLATINE 2055
 QY 1237 --TPGFRSLDNGKLLDRIDRHVEGQFYWFYAFIADALITKLKPYEDTVNRINLDGN 1294
 Db 2056 LNTNKLKENLNSCKLLDELQNTWKNTIEEB-----SPKVIEDIKKSYDEINGRI---GN 2108
 QY 1295 TR--SFIVPVITTEQI-----RKMLSYFYQS-- 1319
 Db 2109 TEMDAEAINVLELVKRONKRSWYTSLGKVTNKMSPDKRLVDRQKISQSFSDSMK 2168
 QY 1320 --GGYSLSLSPYNNIDNLVENDTVIDVNVVKNITTESDIQK-----ELIENILS 1373
 Db 2169 LNFNMINKINKINNHFDQIINN---YSLGVEKMMNYINDSNEKXTSVIEDILNLTK 2224
 QY 1374 KLN1-----EDNKILNHNHTINFGDINES-----NREISITFS 1407
 Db 2225 KISINNSNNNNNNNNNNNDIINFINTKENLITNYENLKGKOLENIYKIFPS 2284
 QY 1408 ILEDI-----NIIIEIDLVSQSKILLSGNCMKLIENGSDIOCKI----- 1447
 Db 2285 KLEBIEKSEKPYFISKAFNNVVKQGGKLSNINRMNNIKNIIIEKETILNKNKFTL 2344
 QY 1448 -----DHIGNGEHQKIPYSIDNETKNGFDYKKEGLFTAFESNESIIRNI 1497
 Db 2345 DTIKEFNKAYDNLPHMRDLYLEESKYDVGKLNIIYIEY-----IYQIRKSNKLNINNF 2399
 QY 1498 Y-MPDSNNLIYSSKDLKD-----TRIINKG-----DYKL-----LIGNYF---KDDMK 1537
 Db 2400 YDIRNTGYIMIHETEKLNIDNEYMNKINKNNNTNNTNKEFDITLRNIKNNTYFVNNNHEIK 2459

QY 1538 VLSFTIEDNTIKLN-GVYLDENGVAQILK-FMNNAKSALNTSNIIMNFLESINIK-NI 1594
 Db 2460 KVIDNFMNHLDTTKANPSTFLPQNEKLVLSIKVINDIK---NIMNEMR-KEQVDVSIQT 2515
 QY 1595 FYNNLDNIEFILDNTNFIISGNSISQFELICD-----KDKNIQPYFINFKIKETS 1645
 Db 2516 CTRNLDIELEKI-----NKIKONNAITDDMVENLFSLDKYKNEFLTKDILKRV 2565
 QY 1646 YTLVGNRQNLIVPE-SYHLDDSGNISSTVINFSQKLYGIDRYNVKV 1692
 Db 2566 QDKY--QEMNKIVRTLTNTVNTDNNEN-----KFALKYLSNVNAFINEI 2605

RESULT 15
 Q26216 PRELIMINARY; PRT; 2771 AA.
 AC Q26216;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
 DE Rhostry protein.
 OS Plasmodium yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5861;
 RN [1]
 RP SEQUENCE OF 379-2771 FROM N.A.
 RC STRAIN=YM;
 RX MEDLINE=97077455; PubMed=8920022;
 RA Sinha K.A., Keen J.K., Ogun S.A., Holder A.A.;
 RT "Comparison of two members of a multigene family coding for high-molecular mass thoptry proteins of Plasmodium yoelii.";
 RL Mol. Biochem. Parasitol. 76:329-332(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YM;
 RX MEDLINE=20446215; PubMed=10989154;
 RA Green J.L., Holder A.A.;
 RT "Structure of the E8 gene encoding a high molecular mass rhostry protein of Plasmodium yoelii.";
 RL Mol. Biochem. Parasitol. 110:167-169(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YM;
 RA Holder A.A.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U36927; AA841263.3; -.
 DR FIR; T28676; T28676.
 DR InterPro; IPR006499; ReticulocyteBP.
 DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
 DR SEQUENCE 2771 AA; 325638 MW; C0CCB9AB6E7ACEF36 CRC64;
 SQ
 Query Match 5.5%; Score 479; DB 5; Length 2771;
 Best Local Similarity 19.9%; Pred. No. 7.3e-10;
 Matches 403; Conservative 339; Mismatches 689; Indels 592; Gaps 101;

QY 6 KACLOQVYVVKFRIQEDVVAIINALEEVHN--MSESSVVEKYLKDKDINNTDNLATY 63
 Db 488 KDTMKIILLIIQYM-NEFKGLNDATWKLKNEGISQKVFVNNQIKQKDSYDE----- 541
 QY 64 KESGRNKALKKFK--EYLTMEVL-ELKNSLTFVEKNLHFIIWGGQINDTAINYNQWKD 120
 Db 542 KKEGFESSLELAKWEKKLEIITELKKVNEETVQLDIKIRELIQKID-----IIEEQKI 597
 QY 121 VN-----SDYTVKVFYDSNAPLINTLTKTIVESATNTLTESFRENLDNDFDYNK 170
 Db 598 VNDLKLELNKTKIETEKTEYIKKAV---DLKKEIKD--NVYID---ELAKEPPYQITK 649
 QY 171 FYKRNELIYDQKHFDYVYKQIEE--NPEFI---DNI-----IKTYSNEY 214
 Db 650 YIEKKQEIYNTIKSDPKTIYVGDIEQLYNEMFSVQESNIEHENKTEILTITKIDNIVY 709
 QY 215 SKDLEALNKYIEBSLNKITANNG-----NDIRNLEK----- 245

710 NKIQNMETVSHLKNVETNNKLSSETILDIKIYIYEITNELNKTLEDFKNKEKGLSNK 769
 246 ---FADELI-VRLYNQELVE---RWNLAAASDILRLSMLEKEDGGVYLD-VDLPGIOPDL 297
 770 IDEYAKENQVLYKSNILEIKHYNDQINIDINKEAKQYDQFKEHMKTIPTPENEI 829
 298 FKSINKPDSITNSWEMIKLEAMKYKEVIGPYTSKNPDMLEBVEQVRSFESALSKSDKS 357
 830 SKTINE-----IKIMKDEFUSKNVY-----NDFKVKYKVESEHNFETELTNK- 874
 358 EIFLPDDIKVSPLEVKIAFANNVINQALISLSDSY-----C--SDLVI 400
 875 -IKTEVSDSEIKKYENKFN-DSKSLNETKSEIEBEYQINILTKKVDDYKIVCLNTNELI 932
 401 NOIKNRYKILNDNLNPSINEGTDFTMTKIPFSKLASI----- 438
 933 TNCHNKOTTLKQKLNQNIKITETNSIDKIYTDKFENILTDKXTELETRFTGLSLNNHES 992
 439 -----SHEDNMFM-----IKITNVLKVGFPADVRSINLSG 470
 993 NNEKELTYFYDLKANLGNKNENLYKQFNEKEKXAVEDIKKQV-----DINKIVSNIE 1045
 471 PGVVTGAYQDLFMKONSTNHLPELRNFBPPKTKISQLTBOETSLWSFNQARAKSQ 530
 1046 ITIYTSIYN---INEDTENEIG-----KSEIELNTKVLEKVKANTNL-----NE 1087
 531 FEEYKKGYPFGALGEDDNLDAQNTVL---DKDYVSKILSSMKT-----RNKEYIHYIVQ 583
 1088 IKEKLDYDFQDFGKEKNIKYPDENKINDIDITLQKIDKSIETLFEIKNGSE--NHIDE 1145
 584 LOG--DKISYEASCNLSPKDPYSILYQKNIEGSETAYYYYVADABIKIDK--YRIPYQ 639
 1146 IKQIDKMLKVPKNTMENEDE-----KEIEKKNIEVEX 1179
 640 ISNKNNTLFIHGKSEFNFTDIFANLVDLSASEIETILNLAADISPKYIEINI--- 695
 1180 IDKKNIT-----YKEIDKLLNEISKIEN-DKTSLE-KLKNINLSYK 1219
 696 -LGCNMF-----SYSIAEETYPGKL-LLKIKDRVSLMPSISOD----- 733
 1220 SLG-NLFLOQIDEBKKAETHIKAMEAYIDDNLNKKKSEIEKEMWINMDIKWDIKHEM 1278
 734 -SITVSANQYEV-----RINBEGREILDSHGWIN--KEES-IKOISSK----- 775
 1279 KALNISHDDYKIYHTTSKNHEEKISDIRKSLKIIODFSEESYINDIKKELEKNVLESQ 1338
 776 -----EYIS-----FN-PRENKIIVSKYLHELSTLLQEIRNN-----ANSSDI-- 813
 1339 NNTDINOYLSKIENIYINILKINKI---KKIIDKYKEYTDEIEKNKKINAELSSEKIIIT 1395
 814 DLEKKWMLTECEINVASNIDQIVEGRIEBAKNLTSDSIN-----YIK----- 856
 1396 QLKENSLSKQCKIKITIDNNYSECIKNITNLKTYVNEKNNTYFRNABEYQNVWS 1455
 857 ---NEFKLIESISDSLDLKHQGLD--DSHFISFEDISKTENGFRIRFINKETG-NSIF 910
 1456 LNFANNIEWADTKSYIILNKKQNGTNTDYNIKELKEHKKSNVYK-----DEAGKNTQE 1510
 911 IETEKEIFSEYATHI-----SKEISNIDKTIIDFNVNGKLKVKVNLDAAEHVT----- 958
 1511 IKKNKELFEKYEQEVTVLLNKYAVELKNKFDKT-KNYSEQIIKEIK--DAHNTFTSOAD 1567
 959 -----LNSAPFIQSLIENYTKESLSNLSVAMKVQVYAOQLFSTGLNITD----- 1003
 1568 KSEKKNBEIKNEQRIEDEVAKNNKSKAI-LDIQLSVEPPFKIKFLKIKDLRTKSDDCDX 1626
 1004 -----ASKVVELVSTALDEITDLPTESEGLPIIATIIDGVSLGAALKELSETNDP 1054
 1627 ETKDIEIKISLSDITQSTKLIENKNILNTEKLESUKN--QKNIEDOKKELDEVNSK 1684
 1055 LLRQIEAKIGIMAVNTAASTAIVTSALGIASGFSILLVPLAGISAGISPLVNNELIQL 1114
 1685 I---KNIESNVNQHKQVY-----EIGIVEKI-----NEIA-- 1711

1115 DKATKVIDYFKHISLAETEGATLLDDKIIMPODDVLSEIDFNNSITLKGKCEIWAERG 1174
 1712 -KANK-----DQIESTOKLIPTIKNLISPFKANDLEGIDTKN---LGK---YNTM 1757
 1175 GSGH---TUTDDIDHFFSPPS---ITYRKPWLSIYDVNLNKKKEKIDFSDKDLMLVFNAPN 1227
 1758 NNIVEEFIKSVDLITHYLETVSKPEITYEQ-----IKNKRIITAQNELLTNIKVN 1807
 1228 RVFGYEMGWTPGFRSLDNDGTKLLDRDRHYEG-----QFYWRY-----FAFTADAL 1274
 1808 KAKSY-----LDDIEANEFDRIVTHFKNKLNDVNDKFTNEYSKVNKGFDNLSNS- 1856
 1275 ITKLKPRYEDTVNRINLDGNTSRFIVPVITTE-----QIRKNLSYSFYG 1318
 1857 INNVK-KSTDENLNLINQTKENYANIVSKKYSYKEAEENIFINIPKLANSLNCTIKS 1915
 1319 SGG-----SYSLSLSPY---NMNIDLNLV-----ENDTWIVDVNVVKNITIESDEIQ 1363
 1916 SSGDLPKFNIALPYLDSQKKTTLTFIPSEKTSYTKISDSYNTLLDILKRSQELQ 1975
 1364 KGE-LIENILSKLNIENKILNHNHTINFYGD-I-NESNRFISLTFISLEIDINIIEIDLIV 1421
 1976 KKEQOALNLIPENELLLHDKVQATNELKDTLSLKNKKEQILNKVLLHKSNEINLSCN 2035
 1422 SKSY-KILLSGNCMKLIENSDDIOQKIDHIGFNGEHQKYPYSYIDNETKNGFI-DYSK 1479
 2036 SQNYDTILESSKYDKIKEKSNNEKEKENLGIN-----FDVKAMEEQFNNDIKDIEK 2087
 1480 KEGLFTAEFSNESIIRNIYMPDSNNLIYSSKDLKDIRIINKGVDKLLIGNYFKDDMKVS 1539
 2088 LENNY-----KHSEKDNVNFSEENNTLQSKKCLKELT-----NAFNAEIK-- 2128
 1540 LSFTIEDTNTIKLNGVYLDENGAQILKFMNNAKSA-LNTSNSLMNPLESINIKNIPYNN 1598
 2129 -----KIEDKIEKNGL--INKLIETRXDCMLFTVKT--VETLKIKTTDY-- 2170
 1599 LDPNIEFTLDNTFISGNSIGQPELICDDKXNIQPYFINPKIKETSYTLVGNRQ--NL 1656
 2171 -----TKFITSATKFSKEFLKYIDATSNLNDIN--TLQTKYDLNQINKHVASM 2218
 1657 IVEPSYHLDDSGNI-----SSTVINPSQKLYGIDRYNVKV 1692
 2219 VADAT---NDNNNLIEKEATKTINNLT--LFTIDS--NKI 2254

Search completed: April 1, 2004, 16:44:42
 Job time : 92.75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:34:35 ; Search time 17.5 Seconds
(without alignments)
5058.244 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700

Perfect score: 8675

Sequence: 1 MNLVNKAQLQKMYVFRIQ.....YLYGIDRYNVKVIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141691 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6759.5	77.9	2366	1	TOXB_CLODI
2	4467	51.5	2710	1	TOXA_CLODI
3	395	4.6	2867	1	RBP2_PLAVB
4	360	4.1	2869	1	RBP1_PLAVB
5	312	3.6	1639	1	MSPI_PLAFW
6	311.5	3.6	1726	1	MSPI_PLAF3
7	309.5	3.6	1682	1	MSPI_PLAF3
8	309.5	3.6	1957	1	SPOF_SCHPO
9	309	3.6	1630	1	MSPI_PLAFK
10	307.5	3.5	1701	1	MSPI_PLAFW
11	306.5	3.5	1726	1	MSPI_PLAFW
12	304.5	3.5	1701	1	MSPI_PLAFW
13	302.5	3.5	2136	1	YCF2_MARPO
14	289	3.3	3433	1	UTRO_HUMAN
15	288	3.3	1875	1	MLP1_YEAST
16	288	3.3	2748	1	NUM1_YEAST
17	285.5	3.3	2894	1	YD96_METJA
18	283.5	3.3	1956	1	ATX1_PLAPA
19	278	3.2	2059	1	TEGU_HSV7J
20	278	3.2	5171	1	BPEA_HUMAN
21	274	3.2	2339	1	RPL1_PLAPA
22	273.5	3.2	1928	1	MYSL_YEAST
23	272.5	3.1	6885	1	SNE2_HUMAN
24	270.5	3.1	1892	1	Y835_RICCN
25	269.5	3.1	2167	1	BEM2_YEAST
26	268	3.1	2663	1	GENE_HUMAN
27	267	3.1	1803	1	YJL3_YEAST
28	265.5	3.1	3210	1	CENF_HUMAN
29	265	3.1	4196	1	DYHC_SCHPO
30	263.5	3.0	1727	1	ALM1_SCHPO
31	261.5	3.0	2875	1	RRPL_TSWV1
32	260.5	3.0	1628	1	NAGH_CLOPE
33	257.5	3.0	1162	1	EXEN_CLOBO

34 257.5 3.0 1679 1 Y109_YEAST P40457 saccharomyc
35 257 3.0 1790 1 US01_YEAST P25386 saccharomyc
36 257 3.0 2014 1 YJ07_YEAST P39526 saccharomyc
37 256 3.0 3144 1 VP13_YEAST Q07878 saccharomyc
38 254.5 2.9 3135 1 S230_PLAFO Q08372 plasmodium
39 254.5 2.9 4092 1 DYHC_YEAST P36022 saccharomyc
40 252.5 2.9 8545 1 ANCL_CABEL Q914m4 caenorhabdi
41 250.5 2.9 1295 1 EXAL_CLOBO P10845 clostridium
42 250.5 2.9 1863 1 YH00_YEAST P38737 saccharomyc
43 250 2.9 3343 1 YOG7_CABEL P34616 caenorhabdi
44 250 2.9 4563 1 APB_HUMAN P04114 homo sapien
45 248 2.9 1274 1 BXF_CLOBO P30996 clostridium

ALIGNMENTS

RESULT 1

TOXB_CLODI
ID TOXB_CLODI STANDARD; PRT; 2366 AA.
AC P18177;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Toxin B
GN TOXB OR TCDB
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]_TaxID=1496;
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90326540; PubMed=2374729;
RA Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
RT "Nucleotide sequence of Clostridium difficile toxin B gene."
RL Nucleic Acids Res. 18:4004-4004(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1271-2366 FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=92293124; PubMed=1603068;
RA Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S., Schulze J.,
RT Sauerborn M.;
RL "Comparative sequence analysis of the Clostridium difficile toxins A
and B."
RL Mol. Gen. Genet. 233:260-268(1992).
CC -!- FUNCTION: Cytotoxin.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X531138; CAA37298.1; -
DR EMBL; X92982; CAA63562.1; -
DR EMBL; X60984; CAA43299.1; -
DR PIR; A27636; A27636.
DR PIR; S10317; S10317.
DR InterPro; IPR007577; CW_binding.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF01473; CW_binding_1; 17-
DR Pfam; PF04488; Gly_transf_sug; 1.
KW Toxin.
SQ SEQUENCE 2366 AA; 269709 MW; E1024BD8B8A56ADF CRC64;

Query Match					
77.9%; Score 6759.5; DB 1; Length 2366;					
Best Local Similarity 76.8%; Pred.No.1:6e-250;					
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;					
Qy	1	MNLVNRQAQLQWYVVFRIQDEBYVALNALAEYHNHSESSVBEKYLKDINLNTDNYL	60		
Ddb	1	MSLVNRQLEKMANVFRFQDEBYVAIDLAEYHNSENIVBEKYLKDINSLTDIYI	60		
Qy	61	NTYKSGRNKALKFKXEYLTMEVLKNNLSPTPEKNLHFHWIGGINDTAINYIQWKD	120		
Ddb	61	DTYKSGRNKALKFKXEYLTVEVLKNNLTPVEKNLHFHWIGGINDTAINYIQWKD	120		
Qy	121	VNSDYTVKYVDNAPLINTLKKTIVESATNTLTESFRENLDPEFDYNKFYKRWEIYY	180		
Ddb	121	VNSDYNVPYDSNAFLINTLKKTVVESAINDTLTFEFRENLDPRFDYNKFFRKRWEIYY	180		
Qy	181	DKQHGFIDYKSIOEENPBFIDNIKTLYSNEYSKOLZALNKYIEBSLNKITANNNDI	240		
Ddb	181	DKQNFINNYKAORENPELIIDDIKVTVLSNEYSKEIDELNTWYIEBSLNKITQNSGV	240		
Qy	241	RNLKPFADELVRLYNQELVERWNLAASDILIRISMLKEDGGVYLDVIDLPQDPOLF	300		
Ddb	241	RNPEEPKNGSFNLYEQELVERWNLAASDILIRISALKEIGWYLDVDMPLPGIQPOLF	300		
Qy	301	INKPDSTINTSWEMIKLEAIKMYKEYIPGYTSKNFMDLDEBQVSFSFALSXSXSEIF	360		
Ddb	301	IEKPSVTVDFWMTKLKAIKMYKEYIPBTSEHDFMDLDEBQVSFSVLAKSDXSEIF	360		
Qy	361	LPLDDIKVSPLEVKYIAFNNSVINQALISLSDSYCDLVINOIKRYKILNDNLNPSINE	420		
Ddb	361	SSLGDMEASPLEVKYIAFNNSVGIIINGLIISVKOSYCSNLIIVQIENRYKILNLSNPAISE	420		
Qy	421	GTDENTTTWKIFSDKLASIGNEDNMFMKIITNYLVKGFPADVRSITNLSGPGVYTGA	480		
Ddb	421	DNDENTTNTNFIDSIMAEANADNGRFMMELGKYLAVGFFPDVKTTINLSGPEAVAAAYQD	480		
Qy	481	LLMFKONSTIHLLPELRNPFPPKTKISQLTEOILTSLWSPNQARAKSQEEYKKGYPE	540		
Ddb	481	LLMFKEGSMMTHLIEADLRNFEISKTNISQSTEQEWASLWSPFDARAKAQOEYKRYNFE	540		
Qy	541	GALGEDDLNLOPAQNTVLDKDYVSKTILLSMKTRNKEYITHYIVQLQGDKISYEASCNLF	600		
Ddb	541	GSLEDGNLDFSNIVVDKEYLELEKISSLARSEGRYIHYIVQLQGDKISYEACNLF	600		
Qy	601	DPYSILLYQKNIEGETAYYYYVADAIEIKEDIKYIPYQISNKRNIKLTFTIGHGKSEFNT	660		
Ddb	601	TPYDSVLFOKNIEDSEAYYYNFGDGEIQEIDKYKIPSIISRDPKILFTIGHGKDEFT	660		
Qy	661	DTFANLVDLSUSSBIETILMAKADISPXYTEINLLGCNMFYSYSISABETYPGLKLIKIK	720		
Ddb	661	DIFAGFDVDSLSTIEAADIAXEKDISPKSIEINLLGCNMFYSINVEETYPGLKLLKVK	720		
Qy	721	DRVSELMPISQDSITYSANQYEVRIINECKREILDHGSKWLNKEESIHKDISKEYISIF	780		
Ddb	721	DKISELMPISQDSIIVSANQYEVRIINSEGRRELLDHSGEWLNKEESIHKDISKEYISIF	780		
Qy	781	NPKENKIIVKSKYLHELSTLLQERINNANSDDILEKKVMLTCECINVASNIDRQIVEGR	840		
Ddb	781	NPKENKIIVKSKNLPSTLLQERINNANSDDILEBKVMLTCECINVSNIDTQIVEER	840		
Qy	841	IBEAKNLPSDSINYIKNEFKLIESISDSLYDLKHQGLDDSHFISFDIKTENGFRIRF	900		
Ddb	841	IBEAKNLPSDSINYIKDEFKLIBESISDALCDLQQONELEDSHFISFDISETDEGFSIRF	900		
Qy	901	INKETGNSIFETEKEIFSEVATHISKEISNIKDTTFDNVNGKLVKKNLDAAEHVNTLN	960		
Ddb	901	INKETGESIFVETEXTIFSEYANHITEESISKIGTIFDVNGKLVKKNLDTTHEVNTLN	960		
Qy	961	SAFFIQSITIENTTKESLNSLVAMKVQVYAQLFSTGLNTITDASKWELVSTALDETID	1020		
Ddb	961	AAPFIQSIEIENSSESLSNVAMKVQVYAQLFSTGLNTITDAKVELVSTALDETID	1020		
Qy	1021	LLPRTLSEGLPIIATIIDGVSIGAAIKESELSTNDPLIRQEIEAKIGIMAVNFTAATAIVT	1080		

Db	1021	LLPTLSEGLPIIATIDGVSIGAAIKELSETSDPLLRQIEAKIGIMAVNLTAITAIT	1080
Qy	1081	SALGTASGFSILLVPLAGISAGIPSLVNNELLQDKATKVIDYFKHISLAETEGAFITLD	1140
Db	1081	SSLIGTASGFSILLVPLAGISAGIPSLVNNELLVLRDKATKVDYFKHVSIVETEGVFTLLD	1140
Qy	1141	DKIIMPQDVLVLSEIDFNNSITLKGCELRWRAEGSGHITLDDIDHFFSPSPITVRKPKWL	1200
Db	1141	DKIIMPQDVLVLSEIDFNNSITLKGCELRWRAEGSGHITLDDIDHFFSPSPITVRKPKWL	1200
Qy	1201	SIYDVINIKKEKIDFSLKOLMWLPNAPNRVFGYEMGWTGFRSLDNDGTKLDRIDRHYEG	1260
Db	1201	SIYDVLEVQKEELDSLKOLMWLPNAPNRVFAWETGWTGFRSLDNDGTKLDRIDRHYEG	1260
Qy	1261	QFYWRYPARIADALITKAPRYEDTNVRINLGNVTRSFIVPVITTEQIRKNSLSYFSGS	1320
Db	1261	EFYWRYPARIADALITKAPRYEDTNVRINLGNVTRSFIVPVITTEQIRKNSLSYFSGS	1320
Qy	1321	GSYSLSLSPYNNIDNLVENDTWIDVDNVKNITIEBDEIQKGLIENILSKLINIEDN	1380
Db	1321	GYAUSLSQYNNMGINIELSESDWIIIDVNVDRVTIESDKIKKGLDIEGLISTLSEEN	1380
Qy	1381	KIILNHTTINFYGDINESNRFSLTFSILLEDINIIIEIDLVSKEYKILLSGNCMKLIENS	1440
Db	1381	KIILNSHEINFSGEYNGSNGFVSLTFSILEGINAIEVLLSKSYKILLSGELKILMLNS	1440
Qy	1441	SDIQOKIDHIGFNGEHOQKIPYSYIDNETKYNGFDIDYKKEGLFTAFENESILRNINYP	1500
Db	1441	NHIQOKIDYIGFNSLQKNIPIYSFVDSGKNGFNGSTKEGLFVSELPDVVLISKVYMD	1500
Qy	1501	DSNNLIYSKOLKDIRIINKGDVKLLIGNYFKDKMKVLSPTIEDTNTIKINGVYLDEN	1560
Db	1501	DSKPSFGYSNNLKDKVKYITKONVNILTVYLUKDDIKISLSLTLQDEKTIKLSNVHDES	1560
Qy	1561	GVAOILKFWNNAKSALNTSGLMNFLESINIKNIIFYNNLDPNIEFILDNTFIISGNSIG	1620
Db	1561	GVAEILKFWNR-KCNTNTSDLSMFLSNKXISIFVNFQLQSNIKFILDANFIISGTTISG	1619
Qy	1621	QFELICDCKQKHOPYINFPIKIBETSITLYVGNRQNLIVBPSYHLDSDGNSISSTVINFSQK	1680
Db	1620	QFEPICDENDNIQPYTIFKNTLETNTLYVGNRQNLIVBPNYDLDSDGDISSTVINFSQK	1679
Qy	1681	YLYGIDRYNKVLIAPNLYT 1700	
Db	1680	YLYGIDSCVKNKVISPNYIT 1699	
RESULT 2			
TOXA_CLODI			
ID	TOXA_CLODI	STANDARD;	PRT; 2710 AA.
AC	P16154;		
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Toxin A.		
GN	TOXA OR TCDA.		
OS	Clostridium difficile.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1496;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VPI 10463;		
RC	MEDLINE=90221894; PubMed=21093310;		
EX	Strain=VPI 10463;		
EX	MEDLINE=90221894; PubMed=21093310;		
EA	Sauerborn M., von Eichel-Streiber C.;		
RT	"Nucleotide sequence of Clostridium difficile toxin A.";		
RL	Nucleic Acids Res. 18:1629-1630(1990).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VPI 10463;		
RC	MEDLINE=90129305; PubMed=2105276;		
EX	Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Lyster D.M.,		

[illegible]

Wilkins T.W., Johnson J.L.;
Molecular characterization of the Clostridium difficile toxin A gene."
Infect., Immun. 58:480-488 (1990).
[3]
SEQUENCE FROM N.A.
STRAIN-VPI 10463;
von Eichel-Schreiber C.;
submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
!- FUNCTION: Only after the enteral delivery of the enterotoxin A may
the characteristic disease called pseudomembranous colitis be induced.
CC CC
CC CC !- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE DIFFERENT OLIGOPEPTIDES.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).

EMBL; X51797; CRA36094.1; --
EMBL; M30307; AAA23283.1; --
EMBL; X92982; CRA63564.1; --
InterPro; IPRO02479; CW binding.
DR InterPro; IPR007577; Gly transf_sug.
Pfam; PF01473; CW_binding_1; 28--
Pfam; PF04486; Gly_transf_sug; 1.
Toxin; Enterotoxin.
KW DRI
SQ SEQUENCE 2710 AA; 308052 MW; OA6E52CEB4C1A421 CRC64;

Query Match 51.5%; Score 4467; DB 1; Length 2710;
Best Local Similarity 50.7%; Pred. No. 5.6e-163;
Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;

QY 1 MLNVKAAQLQMVKVVFRIQDEYVAIINALAELEYHNMSESVVEKYKLKDINNLTNYL 60
DB 1 MSLSKEELIKLAY-SIRPRENEYKITILTNDDEVNKLITNNENKYQLKKLNESIDVPF 59
QY 61 NTYKKSGRNKALKPKKEYLTWVELLEKNSLTPVERNLHPFIWGGOINDTAINYNQWD 120
DB 60 NKYTSSRNRLSNLKDKILKAIVLLINKNSTSPVERNLFHFVVIGGVSDIALEYIQWAD 119
QY 121 VNSDYTKVFDVSNAFLNTLIKTVISATNWTLSPRENLDNPEDYNKYRKMELY 180
DB 120 INAYNIKIWLVDSEAFLTUKKAIVSSTTEAQULLEEIOQPDDNWKFKGMWEFY 179
QY 181 DRQKFIDYKSOIEENFEFTIDIHKTYLSNEYSKDLEALNKYIESLNKITANGNDI 240
DB 180 DRQKFIFYKSIQNKPVTVIDDIISKLVSEYNRDETVLERYTNLSLRKINSNHGDI 239
QY 241 RNLEKFADEDLVRLYNQBELVERWNLAASDIRLSMKEDGGVYLVDVLPGLQPDLPFS 300
DB 240 RANSLTEQEELLANIYSQELLNRGNLAAAASDIVRLALLKNFGGYLVDMPLPGIHSDLFKT 299
QY 301 INKPSITTWSHEMI KLBAMIKYEYPGYTSKNFMDLDEVOBSRESALSCKSKSEIF 360
DB 300 ISRPPSIGLDRWEMI KLBAMIKYKYYINNYTSFNDFKLDOOLKDNFKLIIESKSEKSEIF 359
QY 361 LPIDDIKVSPLEVKIAFANNVINCALISLKDSYCSDLVINOIKRYRKILNDMLNPSINE 420
DB 360 SKLENLVSDLEIKIPALGSVINQALLISKOGSYLTNLVIEQVKNRYQFNLQHLPAIN 419
QY 421 GTDFNTMTKIFSDKLASINENDNMFMFIKITNYLKVGFAPDVIRSTINLGPGVYGAYD 480
DB 420 DNNFDTTTKIFHDSLFSNGATAENSWFRTKAPIYLOVGFMPEARSTILLSGPAGAYSAYD 479
QY 481 LLMFXDNSTNIHLPEELRNPFEPKTKitSQTBGITSLWSFNQAARKSGOFEYKKGYTE 540
DB 480 FINLOENTITEKTIKASDIIIEFFFPENNISQITEQINSLMSFOQASKYOGEFYKRVDTGT 539

QY 1602 NIEPILDTNFIISGNSIGOFELICDKKNIQYFNFINKIKETSITYLVGNRONLIVEPS 1661
DB 1599 NINVIDKYFTLVGKTGLVYVEFCDDNNKNDIDYFGWKTSSSKSTIFSGNGRNVVVEPI 1658
QY 1662 YHLDSDGNSISTVNFOSKYLIGIDRVYVNVKVIAPLNT 1700
DB 1659 YN-PDTGEDISTSLDFSEYELYGIDRVYVNVKVIAPLNT 1696
RESULT 3
ID_RBP2_PLAVB STANDARD; PRT; 2867 AA.
AC Q00799; Q9N2M3;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Reticulocyte binding protein 2 precursor (PvRBP-2).
GN RBP-2 OR RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.
RX MEDLINE=20299192; PubMed=10838229;
RA Galinski M.R., Xu M., Barnwell J.W.;
RT "Plasmodium vivax reticulocyte binding protein-2 (PvRBP-2) shares
RT structural features with PvRBP-1 and the Plasmodium yoelii 235 kDa
RT rhoptry protein family";
RL Mol. Biochem. Parasitol. 108:257-262(2000).
RN [2]
RP SEQUENCE OF 1189-2439 FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites";
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to
CC human reticulocyte cells.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC BMBL; AF184623; AAF76525.1; -.
DR HSP; P03069; IGCM. Signal; Transmembrane; Repeat.
KW Malaria; Receptor; Signal; Transmembrane; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 2867
FT DOMAIN 22 2805
FT TRANSMEM 2806 2826
FT DOMAIN 2827 2867
FT DOMAIN 44 133
FT DOMAIN 560 758
FT DOMAIN 1112 1285
FT DOMAIN 2758 2785
FT REPEAT 2758 2761
FT REPEAT 2762 2765
FT REPEAT 2766 2769
FT REPEAT 2770 2773
FT REPEAT 2774 2777
FT REPEAT 2778 2781
FT REPEAT 2782 2785
SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFFD3 CRC64;
Query Match 4.6%; Score 395; DB 1; Length 2867;
Best Local Similarity 19.1%; Pred. No. 1.4e-07;
Matches 366; Conservative 336; Mismatches 684; Indels 538; Gaps 87;

QY 2 NLV-NKAOLQWVYKFR-----IOEDYVAILNALAEYHNMS----- 40
DB 977 MIKKNADLIKKLOHQTEIEKTFIENEEMSPILLSVIKKEKNEVESDMSEELIKQLNTK 1036
QY 41 --SVVEKYLKLD-----INNLDNYLNTYKSGRKNAL 72
DB 1037 INAILLEYNKSCKDRFNGDDETNLEELDDFKQCOAQOEIKKLTTNY--NVLDNGINVII 1094
QY 73 KKFKEYLTMEVLELKNNSLTVPKLNLFHFIWGGINDTA---INYINQWK-----DVNS 123
DB 1095 KEQHE---KVIILSENHITKDK-----KINEKIQONVNSLNMETKJGLLKINE 1141
QY 124 DYTVKVFYDSNAFLINTLKKTIVE-----SATNNTLESFRE--NLNDPEPDYN 169
DB 1142 D--IKNSRDT-----TIKSKIQEFKKVQTFISGIDVANKKIDAIKKEHVDNDEFDKE 1193
QY 170 KF-----YKRMELIYDKQKHFI--DYKSKQIBENPEFIIDNIIKITYLSNEYSKLEAL 221
DB 1194 KYKDTSPDEKKKSIKAYKVGNTLKELEKWDDEKNIKEVEEAQIQYKRIFIDHDVNL 1253
QY 222 NKYIEESLNKITTANNNGDIRNLEKPADBLVRLNQEL-----VE 261
DB 1254 NDEVEKS--KIV-----MEK-----IELYKKEIDEIKQKTNEYKQDTSNFFYTE 1296
QY 262 RNILAAAS-----DILRISMLKE--DGYVLDVDILPGIQDLPFKSINKPDSITTSWEM- 314
DB 1297 QYNSATQSKAKIEQFINATTKGTSDTSQDINELESIKEEVHKNLQVLUVQESNMEEMR 1356
QY 315 -----IKLEAMKYKEYIPGVTSKN-----FDMLEBEOVSFESALSSKSD 355
DB 1357 KQILSKMOLLILNSETIAKESNNTQNALGFRENKTKLNKTDELLORVAAMTEAAK 1416
QY 356 KSEIFLPDDIKVSPLEVKIAPANNVINQA-----LISLKD--SYCSDILVINOIKNRY 407
DB 1417 KKNIDIALEDAQIDTEVSKIEQINREINWKKDEIKSYLSEIKKEYKDKCTTEISNKRGD 1476
QY 408 KI-----LMDNLNPSINEGTF-----NTWKLPSPDKLASIS 439
DB 1477 KIEFLKFKPNEESNKNVINEINENIRNSQYLDIEDAEKQASTKVFPHKHETTIS 1536
QY 440 N--EDNMFMKITYLVKGFAPDVRSTNLSPGVYT--GAYQDLMFKONSTNIHLE 495
DB 1537 NTFKSEILGVETKQKINKAEIDIMKIERHNSIQVQKGFQENLNKLNPHNYDNAR 1596
QY 496 PELRPFEPKTKISOLTEQIEITSLWSFNOARAKSOFEEYKKGYPGALGE---DDNLDF 552
DB 1597 DELNDKDSNAKV--LIETNLSV-----KXNLSEITNIKQ-----GEKIYSKAKDM 1643
QY 553 QNTVLDKDYVSKILSSMKTRNKEYIHIVQ--QGDKISYEASCNLFSPKDPYSSI--LYQKN 611
DB 1644 QKIKATSENTAETKLEKVKDQSNVNYL-----NQITERNLIVTEKNRLNGIDSTITN 1698
QY 612 IEGS--ETAYVYVADAE--IKEIDKYRIPYQISNRYNKNIKLTFIGHKGEFNT-----DT 662
DB 1699 IEGALKESGNYEIGFLEKLEIEGKRNKLVDTIKTSINST--VGNFSSLFNFDLNQYDF 1757
QY 663 FANLD-----VDSLSEIETIINLAKADI-----SPKYI-----EINLLG-- 697
DB 1758 NRKINDYENKGEIYNEFEGSLNKISENDRASNTSDYNSAKTLERLAEQKVNLLNKE 1817
QY 698 --CNMFYSISISAEETYPGKLLKIKDRVSELMPISQDSITYSANQYEVRIEENEGREIL 755
DB 1818 EANKYLRDVKKVESP--RFIFNMKESLDKINEMIKKEQLTYNEGHGNVQOLVENIKELV 1875
QY 756 DHSGKWKINKEESIIDISSKEYISFNPKENKILVSKYKYLHEUILLQELIRNNANSSDDL 815
DB 1876 DE-----NNLSDLKQATGKN-----BEIQKITHSTLNKAKTILGHVDTSAKYVGIKI 1924
QY 816 EKKWMLTE-----CEINVASNIDRQIVEGRIEAKNLTSDSINYIKNEFKLIESISD---- 867
DB 1925 TPALATTELLGDAKLTAQELKPFESKNVYVLETNMSKNT-----NELDVHKNIOAYKV 1979
QY 868 SLVDLKHQGLDDSHFISFEDISK--TENG----FRIRFINKETGNSIFTETEKEIFSEYA 922

QY	498	LRNFERPKTKI	SOLTBOEITSLWSFNQARAKASFEEVKKGYFEGALGEDNDLDPACN	---	554
Db	1470	VKKIEGIKREILKWKESALT	-FWEESE-----KPKQWCSSHMENAEKGKKIETYLKONGD	1523	
QY	555	-----TVLDK-----	DYVSKKILSSMKTRNKYEIHYIVLOQDKISYBASC-----	595	
Db	1524	GGRKANITDSQMEFVGNYSK	-----AEHAFTTVEAQVDKT--KAFCESIVAVVTK	1571	
QY	596	--NLPKDPYSSILYQKNIEGSETAYYYYVADABEIKDKYR:PYOIS-NKRN	-----K	647	
Db	1572	MDNLFNLSMKVKVCKEKNDAEAKY	-----SAKLAPYDG-RIKARVSENERKISLKK	1626	
QY	648	LTFIGHGKSEFN-TDFFANLDVDSLSSEIETILNLAKADISPKYIEINILGCMNFSYSIS	706		
Db	1627	AKVEKSESSQNDVSTKSLQINDCROQLDSVL	-----NIGRVQNALQYFDS	1675	
QY	707	AEETYPQGLLKIKDRVSELMPISODSITVSANQVEVRINEEGKREILDHSGKWINKEE	766		
Db	1676	ADKSMKSVL-----PISLGAEKSLDKVKAASEYEKNL	-----ETVQNMESNRINVEE	1723	
QY	767	SIKDSSKSYISPNPKENKIIVKSXLHLSTLLOEIRNANSSDIDLEKKWMLTECEI	826		
Db	1724	GSJTDIDDK-----ITDIENLLKMKQYEE--GLLOKIKENADKRSNFP	-----LVGSEI	1773	
QY	827	NVASNIDROI-VEGRIEAKNLSDSINY	-----IKNEFKLTESDSDLYDLKHONGLD	879	
Db	1774	NALLDPSTSIPIKLKKE-YDMTGDLKNYGVQNEIHGEF	-----TKSYNLIETHLSNATD	1828	
QY	880	DSHFISPE-----DISKTENG-----	FRIRFINKETGNSIFITEKEIFSEYA-	922	
Db	1829	YS--VTFEKAQSLFELAKEEHLRRREBEAIFILNDIKKVESILKLKEMMKVSAEYEG	1886		
QY	923	-----THISKEISNIKOTIFDNVNG-----	KLKVKV--NLDAAEHVN-	957	
Db	1887	MKRDHTSVSOLVQDMK-TIVDELKTLNDISECSSVLNNVVSIVKVKKESHADYRRDRANS	1945		
QY	958	-----TLNSAFF-----	IOSLREYNTTKESLSNLSVAMKVQVYAQLFGLSTGLATITDASK	1006	
Db	1946	MYESMVTLANVFLSDEAKISGMEFNA--EMKSNFKTDLELEFP	-----SVINSNSE	1995	
QY	1007	VVELVSTALDEITDLPPLTSEGLPIIATIIDGVSGLAAIKELGETNDPLLRQIEAKIGI	1066		
Db	1996	LLKKIEQDSNDVIQ-KERESEQLAKATDIYV	-----IKLKNEFNKVL-EBAKQKEEV	2047	
QY	1067	MAVNLTRAASTAIVTSALGIAGSPILLVPLAGISAGIPSLVNNELIQLDQATKVIDYFKH	1126		
Db	2048	VSEXVREALKEL-SQVEGIRCHFE-----	NPHRLIDN-----TEELENLKK	2087	
QY	1127	ISLAETEGAFLLDDKLIIMPQDVLVLSEIDFN--NNSITL	-----GKCEIWEAEG	1174	
Db	2088	M-----VTIYKDKKSERESGLOEMENMTYNSITQLEGIVSAGSKEDIKLER	2139		
QY	1175	GSQH-----TLTDDIDHFFSPSITYR-----	KPMLSIDY-VLNTKKEKIDFSKD	1218	
Db	2140	SNEEMRNISEKISITDSKVIEMNSTIDELYLKGNCOAHMISLSIYTANMKTSC	---KL	2195	
QY	1219	LMVLPNAPNRVFGEMGWTPGFRSLDNDGTKLLDRIDRHDYEGQYWPYAFIADALLTKL	1278		
Db	2196	IMINKERENT-----	ERCDVYIKDNSS-----TDGYVETL	2226	
QY	1279	KPYEDTIVRINLDGNRTSFIVPVTITTEIQIKNLISYSPGSGGSYSLSLSPYNNIDNLN	1338		
Db	2227	K-----	GFYGSKULTFSSA-----SEI	2242	
QY	1339	VEN-DTWVIDVNVKNITIBESDTQKGELIENILSKLINBDNKIILNNHTINFYGDINE	1397		
Db	2243	VQNDATYSVNFPAKHEKESLNAIRDIKKELYLFHQNSDISIVEGGV--QNMALYDKLINE	2299		
QY	1398	SNRFISLFTSILEDINIIIFIDLVSKEYILLSGNCKMLIENSDDI--QOKTD-HIGFNGE	1455		
Db	2300	EKR-----EMDELVRN---	ISETKLQMEHSTDVFPKMLIELHKGMMET	2339	

RESULT 5

```

RES001 3
MSPI1 PLAFW STANDARD; PRT; 1639 AA.
ID AC P04933;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5848;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=66014355; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.I.V., Schwarz R.T.,
RA Freeman R.R.;
PT "Primary structure of the precursor to the three major surface
PT antigens of Plasmodium falciparum merozoites.";
RT Nature 317:270-273(1985).
[2]
RN
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC kDa and 19 kDa antigens which are the major surface antigens of
CC merozoites. The maturation take place during schizont.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL; X02919; CAA26676.1; -.
DR PIR; A24594; A24594.
DR PIR; S05603; S05603.
DR PDB; 1CEJ; 28-MAY-99.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor; 3D-structure.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT	CARBOHYD	764	764	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	768	768	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	783	783	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	844	844	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	920	920	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1058	1058	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1165	1165	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1174	1174	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1445	1445	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1526	1526	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	1639	AA; 187618 MW; 2C255B6616C87F6E CRC64;		
Query Match 3.6%; Score 312; DB 1; Length 1639;					
Best Local Similarity 19.5%; Pred. No. 9.9e-05;					
Matches 352; Conservative 282; Mismatches 551; Indels 618; Gaps 94;					
QY	36	NHSESSVVEKYLKLDINNTNYLNTYKSGRNKALKKFKFYLWVLELKNNSLTTPVE	95		
DB	112	NPSDSSSDSAKSYADLKHVRNYLT	159		
QY	96	KHLHFTWIGQINDTAINY-INQWQVNSDYTVKVPYDSNAFLINTLTKTIIVESATNNTL	154		
DB	160	-NTH-----GFKYLIDGYEIN-----ELLYKLN-FYFDLLRAKLNDVCANDYC	201		
QY	155	E-SFRENLDPEFYNNK-----FYRKMEIITYQKHFIIDYKSOIEBENPEFIIDNIITY	209		
DB	202	QIFPNLKIRANELDVLKLVFGYRKPLDNKONVGVKVEDYIKKN-----	245		
QY	210	LSNEYKDLBALNKYIEESLANKITANGNDIRNLEK-----FADEDLVRLYNOELVERNML	265		
DB	246	-----KXTIENINELIEES-KKTIIDKNQKATKEEEKKGYOQAYDL-SYNNKOLEBAHNL	298		
QY	266	AAASDILIRISMLKEDGGVLDVILPGIO-PDLFKSINKPDSITNTSWEMIKLEALIMKYK	324		
DB	299	ISVLE-KRITLKKNEINIKELLKINEIKNPPANGVNTNLT-----LDKNKKIE	348		
QY	325	EYIPGYTSKFNMDLDEVRQSFESALSSKSDKSEIFPLDDIKVSPLEVY--IAPANNVS	382		
DB	349	EH-----EKEI-----KEIAKTIFNIDSLTDPLELEYLREKNKI	386		
QY	383	INQALISLKSYSCLVINOIKN--RYKILNDNLNPSINEGTDFNTMTKIFSKLASIN	440		
DB	387	DISAKVETKES-----TEPNEYPNGVTYPLSYNDINNALNELNSFG-----	427		
QY	441	EDNMFMKITYNLKVGFPADVSTINLSGPGYVTGAYQYDOLLFPKONSTNIHLEPEL	500		
DB	428	-----DLNPFDDYT-----KGFSKNIY-TDNERKK	451		
QY	501	PEPFTKISQLTEQ-EITSLWSFNQARAKSQFEEYKGYPEGALGEDDNDLDPQNTVLDK	559		
DB	452	F-----INEIKKIKI-----EKKIESDKASYEDRSKSLND-----ITK	486		
QY	560	DYVSKILSSM-----KTRNKEYIHYIVLOQDKISYEASCNLFSDKDPYSSILYQK-NIEGS	615		
DB	487	EY--EKLNIYDSKFNNDILNTFEXMMGKRYKYVE-KLTHNTTASVENSCKHLE--	541		
QY	616	ETAYYYYVADAEIKEDIRIPYQISNKNRINKLTPIGHGKSBEFTWTFANLVDVDSLSSEI	675		
DB	542	-----KLTKALKYMEDYSL-----RNIV-----EKELKYKNL-ISKIENEI	578		
QY	676	ETILNLAKADISPKYIEINLLGCNMFPSISAEETYPCKLLKTKDRVSELMPSISQDSI	735		
DB	579	ETLVENIKKDEE-----QLFEKKITKDNKPKDEKILEVSDIV-----	615		
QY	736	TVSANQYEVRIIEEGKEILDHSGWKINKEESIITKDISKEYI---SFNPKENK-----I	787		
DB	616	-----KVQV-----QKVLNKNKDELKKTQILINVELKNINHVPSYKQENKQEPYIL	664		
QY	788	IVKSKYLHELSTLLOEIRNANSSDIDLEKKVMLTECEINVASNIDRQIVEGRI-----	841		
DB	665	IVLKEIDKLKVPKPKVESLINE-----EKNKINTEGQ-----SDNSEPSTEGETGQATT	715		

QY	842	-----EPAKNLTSDSINYIKNEFKLIE-SISDSLYDLKHQ-----NGLDSDSHFISPED	888		
DB	716	KPGQAGSALLEGDSVQAQAEQKQAPPPVPVPEAKAQVPTPPAPVN-----	763		
QY	889	ISTENGFRIRFINK--ETGNSFIPITEKEIIESEYATHISKEISNIKIDIFONVANGKLVK	946		
DB	764	-NKTENVSKLDYLEKLYEFLNTSYI-CHKYILVSHST-----MNEKILK	805		
QY	947	KVNLDAHEVNTLNSAFFIQSLIETNTTKESLNSLNVAMKVQVYLAQFSTGLNTITDASK	1006		
DB	806	-----QYKITREESKJSSC-----	820		
QY	1007	VVELVSTALDETIDLLFTLSEGLPIIATIDGV--SLGAAIKELSE-----TND	1053		
DB	821	-----DPLDLLFNIGNIPVMVSMFDSLNSLSQLFMEIYEKEMVCNLYKLKND	870		
QY	1054	PLIRQETEAIGIMAVNLTAATAITVTSALCIAGSIFILLVPLAGISAGIPSL-VANELI	1112		
DB	871	KIKNLLFEAK-----KVSIVKTL-----SSMOPLSLTPQDKPEVSANDDT	914		
QY	1113	LQDKATKVIDYFKHI-SLAETEGAF-TLLDDKIIMPQDDLVLSID-FNNSITTLGKCEI	1169		
DB	915	HSTWLNNSLKLFEINLSLGNKNKIYQELIGCKSENFYEKILKDSDTFYNESET---NF	970		
QY	1170	WRAEGGGHTLTDIDHFFSSPSITYRKPLWSI-YDVINIKKEXID--FS-----	1216		
DB	971	VKSADDINSLNDSEKKELEEDINKLKTQLSFDLYNKVYKLBRLFPKKTGVGYKM	1030		
QY	1217	-KDLMLVP-----NAPNRVF-----GYEWGTPGPRSIDNGTKLLDR	1253		
DB	1031	QIKKLTLLKEGLESKLNSLNNPKHLQNFVSFFVKKEABIAETE--NTLEN--TKILLK	1086		
QY	1254	IRDHYEG-----QFYWRYPAPFADALITKLKPRYEDTNRVIRNLG	1293		
DB	1087	-----HYKGLVKYNGESSPLKTLSEETQEDNVASLENFVKLSKLEGLKD--NUNLEK	1140		
QY	1294	NTRSFIVP-----VITTEQIRKNLSYSFYSGGYSLSLSPYNNIDNLN-----	1338		
DB	1141	KKLSYLSGLHLLAEULKEVIRKNNYT-----GNPSENNTDVNNALESYKKFLP	1190		
QY	1339	-----VENDTWI--DVDNVKNITI-----ESD	1360		
DB	1191	EGTDVATVVSSESGDTELEQSQPKPEASTHVGAESNTITTSQNVDDVDDVIIVIFGESE	1250		
QY	1361	E-----IQKGE-----LIENILSKLNI-----DNKILNHNHTINFGD	1394		
DB	1251	EDYDDLQGVVGTGEAVTPSVIDNILSKIENEYEVLYKPLAGVYRSKQLENNVMTNVN	1310		
QY	1395	INESNRFISLTSFISLEIDINIIIEIDLVSYSKILLSGNM-----KLIENS	1440		
DB	1311	VKD-----ILNSRNFKNENFKNVLESDLI--PYKDLTSSNYVVKDPYKFLNKEKROKFLSSY	1365		
QY	1441	SDIQOKID-HIGFNGEHQKIPIYSVIDNETKNGFIDYSKKEGLFTAEFSNESIIRNTYM	1499		
DB	1366	NYIKOSIDTIDNFAND---VLGYKILSE-KYKSDLD-SIKKYINDXKGENEK-----YL	1415		
QY	1500	PDNNL-FYSS-KDLKDIRIINKGDVKLLIGNYFKDDMKVYSL-----FTIEDNTIKL	1552		
DB	1416	PLANIETLYKTVDNKIDLFVHL--EAKVNVYTESKNVEVKIKELNVLKTIQD---KL	1470		
QY	1553	NGVYLDEN--GVAQILKFMNNAKSALNTSGLMNFLESINI-----KNIFYNLDPNTEF	1605		
DB	1471	ADFKNNNFVGIADL-----STDYNNHLLITKFLSTGWFENLAKTVLSNLLDGNLQ	1523		
QY	1606	ILD 1608			
DB	1524	MLN 1526			

RESULT 6
MSPL_PLAFC
ID MSPL_PLAFC
AC P04934;
PRT; 1726 AA.

Db 1150 YNGESSPLKT-----LSEVSIQTDNYANLEKFRVLSKIDGKLNLDNLHLGKKLSFLS 1203
Qy 1174 GSGHTTDDHDFSSPSITYRKPLWLSYDVNLNKEKIDFKSKDMLVLPNANRVFGYE 1233
Db 1204 SGLHLITE-----LKEVI---KKNYKNTGNSP----- 1227
Qy 1234 MGWTPGFRSLDNGCTKLLDRIDRHVEGQFYWRYPARTADALITKL---KPRYEDINVRI 1289
Db 1228 -----SENKKVNEALKSYEN-----FLPEAKVTVTVTPPODPVTPSPLSV 1268
Qy 1290 NLDGNTSRFIVPVITTEBQIRKNSYSFYGGSGYSLSLSPYNNMIDLNLV-----EN 1341
Db 1269 RVSSSGS-----TKETQIPTSGSL--TELQOVQOLQNYDEEDSLVLPFGSESDN 1321
Qy 1342 DTWV-----IDVNVVNKNTIESDEIQ-----KGELENIILS-KLNIED 1379
Db 1322 DEYLDQVVTGBAISVTMDNLSGPEYEDVYILKPLAGVYRSLSKKQIEKNIFFTNLND 1381
Qy 1380 NKIILNHH--TINYPGIDNESN---RFSILTSILEDINIIEIDLVSYSKILLSGNC 1433
Db 1382 ---ILNRLKRRKYFLDVLSDLMQFHXHSSNEYIED-----SFKLNSSEQ 1426
Qy 1434 MKLENSSDIOQKIDH--IGFNGEHQKY---IPYSYIDNETKNGFI----- 1475
Db 1427 NTLKSYKIKESVENDIKPAQEGISYEVKULAKYKDDLESIKVKEEKEKFPSPPTT 1486
Qy 1476 -----DYSKKEGLFTAESNESIIRNIYMPDSNNLFIYSSKDLKDIRINKGD----- 1523
Db 1487 PPSPAKTDEQKESKELPFLTN---IETLY-----NNL-----VNKIDDYLN 1526
Qy 1524 --VKLLIGNYPKDMKVSLEFTIEDTWTI--KLVGVYLDENGVAQILKFWNNKASLNITS 1579
Db 1527 LKAKINDCNVEKDEARHKIT-KLSDLKAIDDKID-LFKNNDPEALKKLJND-----DTK 1579
Qy 1580 NSLWNPLESINIKNIFNNL 1599
Db 1580 KMLGKLLSTGLVQNFNTI 1599

RESULT 7

MSPI_PLAF3 STANDARD; PRT; 1682 AA.
AC P19596; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMWSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5834;
[1]
RN SEQUENCE OF 1-1061 FROM N.A.
RP MEDLINE=89166657; PubMed=3327688;
RA Certa U., Rotman D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
RT precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL EMBO J. 6:4137-4142(1987).
[2]
RN SEQUENCE OF 1032-1682 FROM N.A.
RP MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1.";
RL Exp. Parasitol. 81:47-54(1995).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential)
CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC kDa and 19 kDa antigens which are the major surface antigens of
CC merozoites. The maturation take place during schizont.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M35727; AAA29715.1; -
DR EMBL; Y00087; CAA68280.1; -
DR EMBL; Z53267; CAA84555.1; -
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 422 462 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 881 901 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;
Query Match 3.6%; Score 309.5; DB 1; Length 1682;
Best Local Similarity 19.5%; Pred. No. 0.00013;
Matches 333; Conservative 279; Mismatches 587; Indels 513; Gaps 83;
Qy 36 NMSESSVVEKVLKLDINNLTNDYLTNYKSGRNKALKKFKYELTMEVLEKNNSLTPVE 95
Db 100 NPSDDSDSDAKSYADLKHRYQNVLT-----IKELYPELFDITNHLMLTCD 147
Qy 96 KNLHFIWIGQINDTAINY-INQKDVNSDVTVKVYDSNAFLINTLAKTTIVESATNTLT 154
Db 148 -NIH-----GPKYLDGYEIN-----ELLYKLN-FYDPLRAKNDVCANDYC 189
Qy 155 E-SFRENLDPEFDYNNK---FYKRMBEIIYDKQKHFDIYKQSIQIENPEFIIDNIKTY 209
Db 190 QIPFNLKIRANELDVLKLVGRKPLDFIKDNGVKMEDYIKKN-----KTT 236
Qy 210 LSNYSVKOLEALNKVIBESLNKITANNNDIRNLEK---FADEDLVRLNGELVERWMLA 266
Db 237 IAN-----INELIEGSKTTIDQKNADNEEGKKLYQAYDLF-IYNQLQEAHNL 287
Qy 267 AASDILRISMLKEDGVYLDVLDILPGIOPDLFKINKPDSITNTSWEMIKLEATMKYKEY 326
Db 288 SVLE-KRITDTLKNENIKKLEDDIDKIDAEKPTTGVNQILS-----LRLEKSRHEEK 341
Qy 327 IP-GYTSK-NFDMLDEEVQVSFSAISSKSKSEIFLPLDDIKVSPLEVFIAFANNSVI 383
Db 342 IKEIAKTIKFNIDRLFTD-PLELEYLREKNNKVDVTPKSQDPTKSVQIPVFPENGIVY 400
Qy 384 NQALISLKSDSYCSDLVINQIKNRY-KILNDNLNPSINEGTDFNTMTKIFSCLKASINED 442
Db 401 PLPLTDIHNSLAAD---NDKNVSGDLNMPHTKEKINE-----KIITD-----NKE 442
Qy 443 NNMFMKLTNVLKVGFAPDVASTINLSGPGVYTGAYQDLLMFKDNSTNIHLLEPELRNFE 502
Db 443 KRIFI-----NNIKKQIDL-----EENKIN 462
Qy 503 PPKTKISOLTEQEIITSLWSFQARAKSOFEEVKYKYPGEGALGEDDNLDFAGNTVLDKDYV 562
Db 463 HTYKQNKLLLE-----DYEKSKDYELLELEKFKYEMKNFNPNKND-----VVDKIFS 508
Qy 563 SKKILLSMKTR-NKEYIH-----YTVQGGDKISYEASCNLFSDPYFSIILYQKNIEGSE 616
Db 509 ARITYNVEKQRTNNKFSNNNSVNVNQLKKALS------LEDYS--LRKGISEKD 557

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sb-sib.ch).

EMBL; Z70690; CAA94624.1; --
EMBL; AB027811; BAA87115.1; --
PIR; T38077; T38077.
GeneDB SFombe; SPAC1F3.06g; --
KW Sporulation; Coiled coil.
FT DOMAIN 199 785 COILED COIL (POTENTIAL).
FT DOMAIN 804 1235 COILED COIL (POTENTIAL).
FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).
FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 3.6%; Score 309.5; DB 1; Length 1957;

Best Local Similarity 19.1%; Pred. No. 0.00015;
Matches 358; Conservative 316; Mismatches 656; Indels 545; Gaps 95;

38 SESSVWYKYLKLDIN-----NLTD--NYLNTYKKSGRNKALK-KPEYELTMEVLELKN 89
365 SNSQLEEMVLEKESNTIHSQITDAESKLSSPEQ--ENKSLGSDIEY--QNNLSKDK 420
90 SLTPVEKNLHFIWIGGQINDTAINYNQWQVNSDYTV-----KVFYDSNAFL--- 137
421 MVQVSSQLE-----EARGSLAHATGKLAIEINSEDPQNKIKDFEIQDLRACLNSS 474
138 INTLAKTIV-----ESATNTLESFRENLDNPDYKFKRMEIYDKQKFIIDYKKS 192
475 SNEKESALIDKQDQENLREIQEKQKVSESTQSSLOSQDILNEKKKH--EYVES 532
193 QIEENPEFIDNIITKLYSN--EYSKDLEAL-----NKVIESLNKITANNNDI 240
533 QLNE-----LKGELQTSISEHLSQSLSLAEKAEAAVATNNELESKNSL----- 579
241 RNLEKFADEDIVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQLPDLFS 300
580 -----QTLCNAFQEKLA-----SWQLKENEQNFSSLD-----TSFKK 613
301 INKPDSTNTSWEMIKLEAMKYKEYIPGYTSKNFMDLEEVORSFESALSKSDSEIF 360
614 LNESHQBLENNHQTI---TKQLKD-----TSSKLQQLQLE-RANPEQKESTLSDEN--- 660
361 LPLDDIKVPLEVIAFANNSVINQA-----LISLXDSVCSLD----- 398
661 ---NDRTKLL--KLEESNKLKQEDVDLSLEKNIQTLKE---DLRKEEALRFSKLE 711
399 -----VINOIKRYKILNDNLNFSINEGTDENTTKIPSDKLASINED-----NMMP 446
712 AKNREVIDNLKGHEILEQORNDLHSLSDAKNTNAILSELTK-SSEVDKLTANVET 770
447 MIKITNLKVGFPADVRSTINLSGFGVYTGAYQDMLMFXDNSTNIH-----LLEPELRNF 501
771 LTQDSKAMQSFSTSLVNSYQIS-----NLYHEL-----RDDHVNMQSNNTLLESSEK-- 819
502 BPPKTKISQTEQITSLMSNQARAK-----SQPEYKKGYPGALGED-DNLDPAQNT 555
820 --LKTDCENLTQQNMTLIDNYQKLMKHVNQESKVSSELKE--VNGKLSLDLKNLGRSLNV 875
556 VL-DKDYVSKKILASSMTRKNEYTHYVQLCGDKISYEA---SCNLFSPKDPYSILYQKN 611
876 AISND---QILTQALFSLKQY-----DSLEQESAQLSGSLKLSAEKQLLHTEN 922
612 ----IEGSTAYYYYVADAIEIKDKYRIPYQ--ISNKRNIKLTFIGHGXSEFNTDTFAN 665
923 BELHRLDKLTGKLIKIESKSSDLGKLTARQEISNLKEENS-----OSQAITSVKSK 977
666 LDVD-SLSSEETILNKAIDISPKYIEINLLGCMFYSISABETYPGKLLKIK---D 721
978 LDETLSKSKLEADIEHLKNKVSVEVERNAL-----LASNE-----RLMDLKNNGE 1025

QY 722 RVSELMPSIS-----QDSITVSANOYE--VRINEEGKREILDHSG--KWINK----- 764
DB 1026 NIASLQTEIEKKRAENDDLQSKLSVSESEYENLLLSQTNKSLDEKTNQJUEKNVQK 1085
QY 765 --EESIIKDIDSGKEYISEPNK---ENKLI-----VKSXYLHEL-----STL 800
DB 1086 LLDEKQORVLEBELTSKYGLKEENAIQKDELALRKKSKQHDLCANFVDDLKEKSDA 1145
QY 801 LQEIIRNANSSDIDLEKKVMLTECEINVASNIDRQIVEGRIEEAKNLTSDSYINKBKF 860
DB 1146 LEQLTNEKNELIVSLEQSNNSNEALVE-----ERSDLANRLSDMKKSLSDSNVSVIRS 1200
QY 861 LIESISDSLYDLKQONGLDSDHFSIFEDISKTENGFRIRFINKETGNSIFITETK---EI 917
DB 1201 DLVRVNDLDTLKDK--DSLSTQYSEVQCDRDD-----LLDSLKGCSES 1243
QY 918 PSEYA-----THISKEISNIKOTIFDQVNGKLVKKNLDAHAHEVNTLNSAFFIQL 968
DB 1244 FNKAVSARELCTSEIDVPVSEILDONF-----VFNAGNFSELSRLTVLSLENYLDAF 1297
QY 969 IEYNTTKESLGNLSVAMKVQVYQFLSTGLNTIIDA--SKVELVSTALDETIDLLPTLS 1026
DB 1298 NOVNFKKMELNRL-----TTTDAFTKVADLEKQHEHDDWL----- 1336
QY 1027 EGLPIIATIIDGVSLGAAIKELSETNDPLLRQEIIEAKIGIMAVNLTAATAIVTSALGIA 1086
DB 1337 -----IQRGDLKALKDSEKNFLRKEAE--MTENIHS-----LE 1368
QY 1087 SQFSILLVPLAGISAGIP--SIVNNEL-----ILODKATKVIDYFKHIS 1128
DB 1369 EGKEETKEIEIAELSRLEDNQIATNKLQNQLHNLQEIIRLKEDELVLEKESLIISEESLS 1428
QY 1129 LAETGGAFTLDDKIIMFQ--DDL-----VLSEIDPNNSITILGKCQIEMAEAGSSGHTL 1180
DB 1429 --NORQKESLDDAKNEHMLDDTSRKNSSLMKEIESINSSDDKSFLAGAVEKLGAL-- 1486
QY 1181 TDDIDHFFSPSITYRKEWLSIYDVNLIKKEKIDFSCDLMVLPNAPRVFGYEMGWTPGF 1240
DB 1487 -----QKLHSELSLME--NIKSQLOEAKKIQVDESTI----- 1518
QY 1241 RSLDNDGKGLDRIHDHVEGQYRYFAPFADALITKLPKYEDTNVTRINDGNTRSTIV 1300
DB 1519 QELDHE---ITASKNVEGKLNK-----DSIIRDLSENIQLN--NLLABEKS-AV 1564
QY 1301 PVITTEQIEKILSYS-----FYSGSGSYLSLSPYNN--NIDLNLVEND-----TWV 1345
DB 1565 KKLSTEKSESEILQFNRSLADLEYHKSQVESSELGRSKLKLASTTEELQAEERLSLTRM 1624
QY 1346 IDVDVNVKNITTESDEIQGELIENILSKLNIEONKIILNNHTINFYGDINESNRFISLT 1405
DB 1625 LDLOQVQDLSNID-----SLSEDLRTLSLEDS-----VASLOKECKIKENT 1668
QY 1406 FSILEDINIIEI-----DLVSKSYKIL-----LSGNCMKLIENSSDIQOKIDHI 1450
DB 1669 VESLQVLTSLVQARNAELEDVSRVDKIRRRDORCEHLSGKLKL--HSQLEEQHETF 1725
QY 1451 GFNGHQKQIPIYSYIDNETKNGFIDYSKEGLFTAEPFSNESIIRNITMPSNNLFTVSS 1510
DB 1726 --FRAEQQMTQLGFLKETVK-----KQEKLLKCL----- 1753
QY 1511 KDLDIRIINKGDVKKLLIGNYFKODMKVLSFTIEDTNTIKLVGVYLDENGVAQLKPMN 1570
DB 1754 --NLROEQIIPRSSI-LVYESYIRIEKEII--VLOE-----RLNGLELSQQPKGYGTFP- 1804
QY 1571 NAKSALNTSNGSLMNFLESINIKNFYNNLDNFIPLDTNFIISGNSIGOFELICDKDK 1630
DB 1805 -----FKTNRVEMEVLDs-----FKQOVAKLOFLAGAEFIVKFKEL---EKCAAEK 1849
QY 1631 NTQPTFINFKK-----ETSYTYVG-NRQ-----NLTIV--EPSVHLDDS 1667
DB 1850 EQQATFDNVSEKVENLGKSIIEALYFALNREISFRKSLALSAYHNLLVRDSPKPNDSQ 1909
QY 1668 GNISSTVINFSOKYL 1682

506 -PDKDWD-KIFSARYTYNVEKQRYNNKFSNNNSVNVQKXKALSY-----LEDY 555
604 SLLYQKNIQEGSETAYYYV---ADAEIKEDIKYRIPYQISNKEN--IKLTFIG--HGKS 656
556 S---LRKIGISEKDFNHYTYLTKGEADIKKUTE-----EIKSSENKLEKRFGLTHSAN 607
657 EFNTDTFANLDV-DLSLSEIETILNLAKADISPKYIEINLLGCMFMSYSISAEETYPGK- 714
608 -----ASLEVS DIVKLVQVQVLLIKKIEDLRK-IELFLAKAQL-KDSIHVPNIYKPN 658
715 -----LLKIKDORVSELMPISQ-----DSIT---VSANQVEVRINEEG--KR 752
659 KPEPYLLVLKVEVDKLEFIPKVDMLKQBAVLSSITQPLVAASE-----TTEDGGHST 714
753 EILDHSGKWKNEKSIID-----ISSKEYISFNPENKNIIVKSKYLHELSTL 800
715 HTLSQSGETEVEETVEETVGHVTTTITLPPKEESA--PKVKKVENS----- 763
801 LQEIERNANSIDLEKKVMLTECINVASN-----IDRQVEGRIEAKNLTDSI 852
764 -----IEHKNDSNQALTKVYLKLDLFLTKSVYCHKYILVNSNSMDQKLEVYNLTPEE 820
853 NYIKNEFKL-----IESISDSLYDKHQGLDSDHFSIFEDISKTENGFRIRFINKETGN 907
821 NELKSCDPLDLFNQNNIPAMYSLYDSMNIDLQHL--FFELYQKEMYYLHLKLEENHI 878
908 SIFITEKEI-----FSEYATHIS--KEISNIKOTIFDN-----VNGKLVK 946
879 KKLBEQKQITGTSSTSPGNTVTNTAQSHNSQONQSNASTNTQNGVAVSSGPAV 938
947 KYNLDAAEHVNLTNSAFIQSLIEY-NTTK-ESLSNLSVAMKVQVYALFSTGLNTITD- 1003
939 BESHDPVLVLSISDLKGLVSLNGLNKTKVPNPLTISTTEWKEPYENILKNNDTYFND 998
1004 -----ASKVVE-LVSTALDETIDLLPTLSEGLPIIATIIDGVSL----- 1041
999 IKQVKSNSKVITGTEQKNALNDEIKKIDTLQSLDFLYNKYKLDRLFKNKKELGQ 1058
1042 -GAATKEGETNDPLRQIEAKIGIMAVNLTAASTAVTSALGIASGFSILL--VPLAG 1098
1059 DKMQIKKLT-----LLKEQLESKLSL-----NNPHNVQLQSVFFNKKKEAE 1101
1099 ISAGIPSLVNNELIILQDKATKVIDYFKHISLAETGATLLDDKILMQDDLL-----V 1151
1102 IAEENTENTKILLKHYKGLVYKNGSSPLKT-----SEVSIOETDNYANLEKFA 1155
1152 LSEIDFN-NNSTIGKEIWEAEGSGHTLDDIDHFFSSPITRYKPKWLSYDVLNKK 1210
1156 LSKIDGLNDNLHGLKKLSLSSGLHLITE-----LJ 1189
1211 EKIDFSKDLMLVLPNAPNRVFGEMGTQGRSLDNDGTLLDRIDRHYEGOFYRYFAFI 1270
1190 EVI---KKNVYTGNSP-----SENNKYNKALSKYEN-----FL 1220
1271 ABALITKL-----KPYEDNVRINDGNTRSFIVPITTEQIRKNLSYFVGSGGSYLS 1326
1221 PEAKVTVTVPQDPTPPLSVRVSGSGS-----TKEETQIPTSGSLI--TELQVQV 1273
1327 LSPYNNIDNLV-----ENDTWV-----IDVNVKNIIEISDEIQ----- 1363
1274 LQNYDEEDSLVLPPIFGESDNDEYLDQVVTGEAISTMDNLSGFENEDYVILKPLA 1333
1364 -----KGLIENILS-KLNIEDNKIILNNH--TTFYGDINESN-----RPISTLPSILE 1410
1334 GYVRSKKGQIEKNIIIPNLND---LNSRLKRYKFLVDVLESOLMQFKHLSNNEYIIE 1390
1411 DINIIEIDLVSQKYLKSGNCKMLIENSDDIQKIDH-IGFNGEHQKY---IPYSYID 1466
1391 D-----SFKLLNSEQKNTLLSKYKIKESVENDIKFAQEGISYEKVLAKYKD 1438
1467 NETKNGFI-----DYSKKEGLTAEPFNSIESIRNIYMPDSNNL 1505
1439 DLESIRKVTKEKEKFPSSPPTTPPSPAKTDSQKESKFLPLTN-----IETLY-----NNL 1491

QY 1506 FIYSSKDLKDIRINKGD-----VKLLIGNYFDDMKVLSLSTIEDTNTI--KLVGY 1556
Db 1492 -----VNKIDDYLINLAKINDCNVDEAHVKIT-KLSDLKAIDDKID-LF 1536
QY 1557 LBENGVAQILKPMNNAKSALNTSINMFLBESINIKNIFYNNL 1599
Db 1537 KNTNDPEAIKKLIND-----DTKDMLGKLLSTGLVQNFNPNTI 1574

RESULT 11

MSPI_PLAPP STANDARD; PRT; 1726 AA.
ID MSPI_PLAPP AC P50435;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (GP195)
GN MSP-1.
OS Plasmodium falciparum (isolate Palo Alto / Uganda).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57270;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89005525; PubMed=3049134;
RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA Siddiqui W.A.;
RT "Plasmodium falciparum: gene structure and hydropathy profile of the
RT major merozoite surface antigen (gp195) of the Uganda-Palo Alto
RT isolate";
RL Exp. Parasitol. 67:1-11(1988).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (potential).
CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC kDa and 19 kDa antigens which are the major surface antigens of
CC merozoites. The maturation take place during schizont.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M37213; AAA29611.1; .
DR InterPro; IPR006209; BGF_like.
DR Pfam; PF00008; BGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196174 MW; 5859CEFA2F9A026 CRC64;

Query Match

Best Local Similarity 3.5%; Score 306.5; DB 1; Length 1726;
Matches 349; Conservative 284; Mismatches 602; Indels 525; Gaps 90;

QY 51 DINNLDTNLTNYTKYGRNKALKKFKLEYLTMEVLELKNNSLTPEKKNLHFIWGGQINDT 110

RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RT Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
PLasmodium falciparum merozoites: studies at the genetic level";
RL EMBO J. 4:3823-3829 (1985).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
(Potential).
CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC kDa and 19 kDa antigens which are the major surface antigens of
CC merozoites. The maturation take place during schizont.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X05624; CAA29112.1; -;
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1701
FT CARBOHYD 110 110
FT CARBOHYD 239 239
FT CARBOHYD 470 470
FT CARBOHYD 536 536
FT CARBOHYD 607 607
FT CARBOHYD 802 802
FT CARBOHYD 899 899
FT CARBOHYD 919 919
FT CARBOHYD 965 965
FT CARBOHYD 991 991
FT CARBOHYD 1089 1089
FT CARBOHYD 1196 1196
FT CARBOHYD 1598 1598
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA38 CRC64;

Query Match
Best Local Similarity 3.5%; Score 304.5; DB 1; Length 1701;
Matches 355; Conservative 284; Mismatches 603; Indels 549; Gaps 91;

QY 36 NMSESVVEKYLKLDINLNTYNTYKSGRNKALKKPKFYLTMEVLEKNNSLTPVE 95
DB 106 NPSDSSDSTNTYADLXRVQNYLFT-----IKELKYPELFDLTNEMLT-LS 152

QY 96 KNLHFIWIGQINDAINYNQWVNSDYTKVFDYSDNAPLINTLKKTIVESATNTLE 155
DB 153 KNV-----DGFKYLDIGYEEIN-----ELLYKLN-FYDILLRAKLANDACANSYQ 196

QY 156 -SFRENLDPEFDYK---FVRKEMELIYDKQKHFIDYKSOEENPEFIDNIKTYL 210
DB 197 IPFNLIKIRANEDLVLIKIVGKRPFLDNKQNVGMEDYIKGN-----KTTI 243

QY 211 SNEYSKOLEALNKYTEESLNKLTANNQNDIRNLEK--PAEDSLVRLNCELVERNLAAA 268
DB 244 AN-----INELGSKTIIDQKNADNEEGKKLQAOVNFYINQLOEARNLISV 295

QY 269 SDILRISMLKEDGGYLDVDPGLPGIPLFKSI--NKPDSITNTSWEMIKLEA-IMKYKEY 326
DB 296 LE-KRIDTLKKVENIKKULEDDIKITDAENPTTSGKFNLPENK-----KKE 342

QY 327 IGYTSK-----NFDMLDEEVQSFESALSSKSDKEIFPLPDDIKVSPLEVKIA 376
DB 343 VEGHEEKIEAKTIKFNIDSFTD-PLELEYLREKNKVDVTPKSDQPTKSVQIPKVP 401

QY 377 FANNVINGQALLSLKSDYLINQIKNRY-KIINDNLNPSINEGTFNTMTKIFSDKL 435
DB 402 YPNGIVYPLPLTDIHNLAAD-----NDKNSYGLMNPDTKEKINE-----KIITD-- 447

QY 436 ASISNEDNMFMKITYNLKVGFPADVSTINLSPGVYVTGAYQDMLMKONSTNIHLE 495
DB 448 ----NKRKIFI-----NNIKQIDL----- 464

QY 496 PLRNPFPKTKISOLTQEITSLWFOARAKSOFEBYKKGYPGALGEDNDLPAQNT 555
DB 465 -BEKINHTEONKKLE-----DYKSKDYEELEKFBY- -MKFNNN- 505

QY 556 VLDDKVVSKILLSSMKTRKE--YIH-----YIVLQGDGKISYEASCNLFSKDPY 603
DB 506 -FDKVDVD-KIFSARYTNNVEKQRNNKFNSSNNNSVYVQKLKALS- -LEDY 555

QY 604 SSILYQKNIQEGSETAYVVV---ADAIEIKDKYRIPYQISNKRN--IKLPIG--HGKS 656
DB 556 S---LRKGISEKDFNHYITLTKGLEADIKKLE-----EIKSSENKILEKQKGLTHSAN 607

QY 657 EFNTOFTANLDV-DSLSBEIETILMAKADISPKVIEINLLGCNMFSSISAEETYPGK- 714
DB 608 ----ASLEVSIVLQVQVLLIKKIEDLRK-IELFLKNAQL-KDSIHVPNTYKQFN 658

QY 715 ----LLAKIKDVSSELMPSISQ-----DSIT--VSANOYEVRIINEEG--KR 752
DB 659 KPEPYVILVKKEVDKLEFIPKVKMDLKKOEAVLSSITQPLVAASE-----TTEDGGHST 714

QY 753 EILDHSGKWINKESIIKD-----ISSKEYISFNPKNKIIIVKSKYLHELSTL 800
DB 715 HTLSQSGTEVTEETEVEVTHGHTTTTITLPPKEESA--PKEVKVENS----- 763

QY 801 LOEIRNANSSDIDLEKVMYTEC-----INVA-SNIDROIVEG---RIBEA 844
DB 764 ---IEHSNDNSQALTKTVLAKLDEFLTKSYCHKYLVSNSMDQKLLVYNLTPEE 820

QY 845 KNLTS-----DSINYKNEFKLIESDSL-YDLKHQNGLDSDHSPISFEDISTENGFRIR 899
DB 821 KELKSCDPLDLFNQNNIPAMYSLYDSMNLDQ-----LFFELYQKEMYYLH 870

QY 900 FINKETGNSIFETEKEI-----FSEYATHIS--KEISNIKOTIDN--- 939
DB 871 KUKENHIKLEEOQKITGTSTSPGNTVTNTAQATHSNSQOQSASSTNTQNGVA 930

QY 940 -VNGKLVKVNLDAAHEVNTLNSAFFIQLSIEY-NTTK-ESLSNLSVAMKVQVYQLFT 996
DB 931 VSSGPAVVESEHDPLTVLSISNDLKGIVLLNGLNKTVPNPLTISTTEMEFYEINLKN 990

QY 997 GLNTHD-----ASKVVE-LVSTALDETTDLPTLSEGLPIIATIIDGVS- - 1041
DB 991 NDTYFNDDIKQFVKSNSKVIITGLTETQKNALNDEIKKLDTLQLSFDLYNKYLKLDRL 1050

QY 1042 ----GAAIKELSETNDPLLRQEIHAKTIGIMAVNLTAATAVTSALGIASGFSIL 1092
DB 1051 NKKKELGQDMQIKKLT-----LLKBEQESKNSL-----NNPHNVLQNFVSF 1093

QY 1093 L--VPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETEGAFITLDDDKIIMPQDDL 1150
DB 1094 FNKKKEAEIAETENTLENTKILLKHVKGLVKYNGESSPLKT-----LSEVISIQTEDNY 1147

QY 1151 ----VLSEIDFN-NNSITLGKCEIWRAGEGSGHTLTDIDHFPSSPSITYRKPLWSI 1202
DB 1148 ANLEKPRALSKIDGKLDNLDNLHGLKKLSFLSSGLHLITE----- 1187

QY 1203 YDVNLTKKIDFKIDKDLAVLPNAPNRVFGYMGWTFPFRSLDNDGTGLDLDRIDHYEGOF 1262
DB 1188 ----LKEVI---KNKNTGNSP-----SENNKVKNEALKSYEN-- 1218

QY 1263 YNRYFAFADALITKL-----KRYEDTVNRINLDGNTRSFIVPVITTEIOIRKNLSVFPY 1318
DB 1219 ----FFPEAKVTVTVPQDPVTPSPUSRVSSSGS-----TKETQIPTSGLL- 1265

QY 1319 SGGSYSLSPYNNIDNLV-----ENDTW-----IDVNNVKNITIED 1360
DB 1266 TELQOVVQLQNYDEEDSLVLPITGESDNEIDYLDQVVTGSAISVTMDNLSGNEFYD 1325

366	DB	SEIYIKIEELPSDYIYKFSKYILYEGKSKTIIKQSFNNNIYVYKCLNSINFNTIFYPDS	425
134	QY	NAFLINTLAKTIVESATNWT--LESGFRENLDPEFY-----NKFVRKMEIYIDYQKHF	187
426	DB	NNLLFDWLKKNY--INNPKLKF-----LIYSSISNOF-----ILFFQKQSK	468
188	QY	DYYSQIOENPEILDNIITKYL--NEYSKDL-----EALNKXIEESLNKIT	233
469	DB	SFNKNVLKXNSKDVITNPFNSKENKIEINNFSKIYAFEILSINEINUKFY---INKIS	525
234	QY	ANNGNDIRN---LEKFADEDVRLYNQELVERMN-----LAAASDILRIS-ML	277
526	DB	LKNINKKKQRFVNLKIKSDNPRFINLWKIKYSSQOFVSNNSFLNPAPEILQOYYL	585
278	QY	KEDGGVYLDVILPGIQDPLFKSINKPDS-----INTSWEMIKLEAMK----	322
586	DB	KKKNILF-----FKKLEVPSPFFVFOYKCKLNLIFLKASLEKILKRNK	632
323	QY	-----YKEVTPGYSKNFMDLEEVORSFESALSSKDSSEIFLPLDDIKV-----	368
633	DB	KFTISIKLFKKFYKNKLNENGEYKIESQILONEKELNKKRKNFOFNP--NIKILSFVNS	690
369	QY	SPLLEVIA---PANNSVINOALISLKSDSVLINOIKRYKILNDNLNPSINEGTDPN	425
691	DB	SKNNIYLQNKYFNNKLNINNKLITWK-----KISNKLIVSNSEYN-----	730
426	QY	TTWKIPSDKLASINSDNNMFMIKITNYLKVGFAPDVRSTINLSPGVYTGAYDILLMPK	485
731	DB	-----KI-----IWNKKMKF-----FSPSKNSVL-----DTFFPN	756
486	QY	DNSTNIHLPEPELURNPEFPKTKISQITBEIITLWS-----FNQARAKSQPEYKKGIFE	540
757	DB	KKSFNIITV-----IPDKLKKIQLNFQEOIKILNCFSLFPNSKNIK-KTIFKNSYP-	807
541	QY	GALGEDNDLDAQNTVLDXDYSSKTLSSMKTNRKEYIHVIVLOQGDKTSYBASCNLPSK	600
808	DB	-----INENUTTFSPNDKEFNIFHELPISEINNDFLMRFPFK	846
601	QY	DPYSSILYQKNIEGSETAYVYVADAEI--KEIDKYRIPYQISNKENIKLTFIGHKSPF	658
847	DB	-----YLYRIYKDKELLFNPENRQLLQNFPEKTKI-LTFI-----DF	884
659	QY	NTDTFANLVDVLSSEIETLNLAKADISPKYIEINLGCNMFYSISABEYYPKLLK	718
885	DB	LQD-----PELNANNRFFIHEKKTIKNN---NLL-----YLRLLKIF	919
719	QY	IKDVSSELMPSISQDSITSYANQVYRINEEGKEITLDHSGKWINKESIIKDSSKEYI	778
920	DB	LKDKRNFL-----INE-----IK-----	933
779	QY	SFNPENKIIYVSK-----YLHELSTLQELIRNNANSSDIDLEK	817
934	DB	SFIEKNNLFIKSQLSNVLVKNYSKFFDNI FNFHFLKQKEKNIEIILNNQNVFESLLK	993
818	QY	KWMLTECEINVA-SNIDRQIVEGRIEBAKNLTSDSINYIKNEPKLIESDSIY-----D	871
994	DB	KTYLKNLNLNNSYSKFSKIF--IFOLLNILN-----KNNYKTFQWISLTFYSKLN	1044
872	QY	LKHQGLDDSHFTSPEDISKTENGPR-IRINK-----ETCNSIFETEKEIPSEYATHIS	926
1045	DB	YKIQNKTEKKNYCNKNISYKKKKIKITVNFPEKNLFOITNNSWFFLE---MWEYNTYIL	1101
927	QY	KEISNIKDTIFDNVN-----GKLVKKNLDAAEVNTLNSAFFIQS-----	967
1102	DB	LQI---IQETTFQITDVLEYFKKKIIEK-----NLKFLKSKKISLKTISPHN	1147
968	QY	-----LIEYNTTKESLN-----LSVAMKVOVYAO-LPSTGL	998
1148	DB	FKLXWNLRFNEINYKNKYLLNFWLSDNLINCNLNLYWVIFSLVIFIPLYYQKIESIII	1207
999	QY	NT-----ITDAS-----KVELYSTALDETDILLPLISEGLPIIAT----	1034
1208	DB	GSDCFHLWKNPETIIQYITDASRSIYFPLKTRNKTKANTENLLSYFQNLTHYITNIKF	1267

QY 1035 -----IIDGVSIGRAIKELSETNPDLLROEI--PAKIGIMAVNLTAASTAIPTS 1081
 Db 1268 YLLTKONLKKWLINNKTL-----DLRRKRLVQSLITNKIQNYGFELN-SNKOFFTS 1321
 QY 1082 ALG-----IASGFSILLVPLAGISAGIPSLVNNELIILQDKATKVIDY---- 1123
 Db 1322 YFGQIINQOGLLVFVLAQFFQ-----KNLNNSL---DLANKWIVSFWSH 1365
 QY 1124 ---FKHISLAETEGAFLLDDKIIMPODDLVLSEIDFNNSITL-----GKCEIWRAE 1173
 Db 1366 KIFSSQKLQRTNNI-----ELGFQNIIPVLPQGLSFGKILLIGPIE 1407
 QY 1174 GSGHSLTDDIDHFF-----SSPSITYRKP-----WLSI-YDVNLKKEKIDFSKOLM 1220
 Db 1408 TGRSYLIKNAAESYVFLKISINKLLYNKPDVITESMMNLLIESRLRLNLTLDPAKMM- 1466
 QY 1221 VLPNAPRVFCYEWGWPFFSL-----DNDGTKLLDIRDHVEGOF---YWRYPFA 1268
 Db 1467 ---SPECII-----WQNIHQNLNRLTONVESDPTFLGILLKYFQDPSFKTKNNII 1516
 QY 1269 FIADALITK-----LKPVED--TNVRI-NLDGNTSRFIVPVITTE---QIRKNLSYSF 1316
 Db 1517 VIGSTHLPKKYDPAISNRLDKIINVLFNISOQKKQF--PLLKKKNFQKRENL--PF 1572
 QY 1317 YGSGYSLSLSPYNNMIDLNLVNDTWV-----IDVDNV-----VKNITIEDS 1360
 Db 1573 LNEFGSRMTG---YNLR-DLSALTNELLISITKNRGSFIDTDTLKLAFHRQIFGLTYTNN 1628
 QY 1361 BQKGELENLTSKL--NIENKILNHTNFP--YGDINESNPFSLTFSILE---DIN 1413
 Db 1629 KINFDRIFKIVYKVGKTIION--ILIKSSNNLNLGNFLWKQNFYLSWKYLEPSIDES 1687
 QY 1414 IIIEIDLVSXKIL--LSGCMK-----LIENSDDIQKIDHIGFNGEHQKYPIYSIDN 1467
 Db 1688 IIKELTILT---HILACLAGTAARDSWELKEKKAESLLPIDKL---VENDFTLAFSILES 1741
 QY 1468 -----ETKYNGFIDYSKKEGLTAFTAESNESIIRNIYMPDSNNLFIYSSK 1511
 Db 1742 FSEPPFMLEICOTNVNKKNIIEFTSKSM-----NIMQNGIFAIANKKFIVTON 1793
 QY 1512 DLKDIRINKGVDKLLIGNYFKDDMKVLSPT----- 1543
 Db 1794 HLQ-----YKSSLSQQISFNKKKNVEFKNTSPPRFLRSPFRSLNLF 1836
 QY 1544 -IEDNTYKLVGLDENGVAQLKFMNAKSALNTSNLSNLESLNINIKNIFNNLDPN 1602
 Db 1837 WIKRPNDFFSYKF---GFTKKKGYLFSAN--LQKNYQGFIEKKKKQLLYERILPR 1890
 QY 1603 I-----EFILDTNFIISG 1615
 Db 1891 IRRNVQLESQFEILLEEQFEILG 1916

RESULT 14
 ID UTRO HUMAN STANDARD; PRT; 3433 AA.
 AC P46339;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Utrophin (Dystrophin-related protein 1) (DRP1) (DRP).
 GN UTRN OR DMDL (Human).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93096045; PubMed=1461283;
 RA Tinsley J.M., Blake D.J., Roche A., Fairbrother U., Riss J.,
 RA Byth B.C., Knight A.E., Kendrick-Jones J., Suthers G.K., Love D.R.,
 RA Edwards Y.H., Davies K.E.;

RT "Primary structure of dystrophin-related protein.";
 RL Nature 360:591-593 (1992).
 RN [2]
 RP INTERACTION WITH SNTB1.
 RX MEDLINE=95146543; PubMed=7844150;
 RA Ahn A.H., Kunkel L.M.;
 RT "Syntrophin binds to an alternatively spliced exon of dystrophin.";
 RL J. Cell Biol. 128:363-371 (1995).
 RN [3]
 RP INTERACTION WITH SNTA1 AND SNTB2.
 RX MEDLINE=96162017; PubMed=8576247;
 RA Ahn A.H., Feener C.A., Gussone E., Yoshida M., Ozawa E., Kunkel L.M.;
 RT "The three human syntrophin genes are expressed in diverse tissues,
 RT have distinct chromosomal locations, and each bind to dystrophin and
 RT its relatives.";
 RL J. Biol. Chem. 271:2724-2730 (1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 144-261.
 RX MEDLINE=99141377; PubMed=9887274;
 RA Keep N.H., Norwood F.L.M., Moores C.A., Winder S.J.,
 RA Kendrick-Jones J.;
 RT "The 2.0-A structure of the second calponin homology domain from the
 RT actin-binding region of the dystrophin homologue utrophin.";
 RL J. Mol. Biol. 285:1257-1264 (1999).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 31-256.
 RX MEDLINE=20113481; PubMed=10647184;
 RA Keep N.H., Winder S.J., Moores C.A., Walke S., Norwood F.L.M.,
 RA Kendrick-Jones J.;
 RT "Crystal structure of the actin-binding region of utrophin reveals a
 RT head-to-tail dimer.";
 RL Structure 7:1539-1546 (1999).
 CC -!- FUNCTION: May play a role in anchoring the cytoskeleton to the
 CC plasma membrane (By similarity to dystrophin).
 CC -!- SUBUNIT: Interacts with the syntrophins SNTA1; SNTB1 and SNTB2.
 CC -!- SUBCELLULAR LOCATION: Neuromuscular junction.
 CC -!- TISSUE SPECIFICITY: Muscle.
 CC -!- SIMILARITY: STRONG, TO DYSTROPHIN.
 CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
 CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
 CC ABP-120, ABP-180, OR BETA-PODRIN).
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC -!- SIMILARITY: Contains 1 WW domain.
 CC -!- SIMILARITY: Contains 20 spectrin repeats.
 CC -!- SIMILARITY: Contains 1 ZZ-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X69086; CAA48829.1; -.
 DR PIR; S28381; S28381.
 DR PDB; 1BHD; 16-FEB-99.
 DR PDB; 1QAG; 01-JAN-00.
 DR Genew; HGNC:12635; UTRN.
 DR MIM; 128240; -.
 DR GO; GO:0005856; C:cytoskeleton; TAS.
 DR GO; GO:0005624; C:membrane fraction; TAS.
 DR GO; GO:0005886; C:plasma membrane; TAS.
 DR GO; GO:0006936; P:muscle contraction; TAS.
 DR GO; GO:0007517; P:muscle development; TAS.
 DR InterPro; IPR001589; Actbind actinin.
 DR InterPro; IPR001715; Calponin-like.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR001202; WW_Rsp5_WWP.
 DR InterPro; IPR000433; ZnF_ZZ.
 DR Pfam; PF00307; CH; 2.
 DR Pfam; PF00435; spectrin; 18.
 DR Pfam; PF00397; WW; 1.

DR Pfam; PF00569; ZF; 1.
 DR SMART; SM00033; CH; 2.
 DR SMART; SM00150; SPEC; 19.
 DR SMART; SM00456; WW; 1.
 DR SMART; SM00291; Znf_ZZ; 1.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00020; ACTININ_2; 1.
 DR PROSITE; PS00021; CH; 2.
 DR PROSITE; PS01139; WW_DOMAIN_1; 1.
 DR PROSITE; PS00020; WW_DOMAIN_2; 1.
 DR PROSITE; PS01357; ZF_ZZ_1; 1.
 DR PROSITE; PS0135; ZF_ZZ_2; 1.
 KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
 KW Repeat; 3D-structure; Actin-binding; Zinc-finger.
 FT DOMAIN 1 246 ACTIN-BINDING.
 FT DOMAIN 31 135 CH 1.
 FT DOMAIN 150 252 CH 2.
 FT REPEAT 253 308 SPECTRIN 1.
 FT REPEAT 309 417 SPECTRIN 2.
 FT REPEAT 418 526 SPECTRIN 3.
 FT REPEAT 541 637 SPECTRIN 4.
 FT REPEAT 687 798 SPECTRIN 5.
 FT REPEAT 803 902 SPECTRIN 6.
 FT REPEAT 1016 1083 SPECTRIN 7.
 FT REPEAT 1125 1230 SPECTRIN 8.
 FT REPEAT 1248 1334 SPECTRIN 9.
 FT REPEAT 1432 1541 SPECTRIN 10.
 FT REPEAT 1544 1649 SPECTRIN 11.
 FT REPEAT 1652 1753 SPECTRIN 12.
 FT REPEAT 1910 1968 SPECTRIN 13.
 FT REPEAT 1976 2081 SPECTRIN 14.
 FT REPEAT 2258 2333 SPECTRIN 15.
 FT REPEAT 2399 2440 SPECTRIN 16.
 FT REPEAT 2443 2556 SPECTRIN 17.
 FT REPEAT 2559 2636 SPECTRIN 18.
 FT REPEAT 2658 2688 SPECTRIN 19.
 FT REPEAT 2691 2797 SPECTRIN 20.
 FT DOMAIN 2812 2845 WW.
 FT ZN_FING 3064 3111 ZZ-TYPE.
 FT HELIX 32 46
 FT TURN 47 49
 FT TURN 56 62
 FT HELIX 64 74
 FT TURN 75 75
 FT HELIX 86 102
 FT TURN 103 104
 FT HELIX 112 117
 FT TURN 118 118
 FT HELIX 120 134
 FT TURN 135 135
 FT HELIX 136 149
 FT HELIX 152 163
 FT TURN 164 164
 FT TURN 168 169
 FT HELIX 177 179
 FT TURN 180 181
 FT HELIX 183 191
 FT TURN 192 192
 FT HELIX 194 196
 FT HELIX 199 204
 FT HELIX 207 222
 FT HELIX 230 233
 FT HELIX 240 250
 FT TURN 253 253
 SQ SEQUENCE 3433 AA; 394488 MW; EAE8DB409F859E5B CRC64;
 Query Match 3.38; Score 289; DB 1; Length 3433;
 Best Local Similarity 18.6%; Pred.No. 0.0019;
 Matches 318; Conservative 293; Mismatches 612; Indels 488; Gaps 85;
 QY 27 ILNALLEEYHNMSSESSVEKYLKLDINNLTDTNLTNTYKSGRNKALKKFKYEYLTMEVL-E 85
 DB 599 LLNWLKWTAIQTTEIKENYKMQDTEMKKK-LKALEKEQREIRIPEADELNTGTGILVE 757

QY 86 LKNNSLTPVEKNLHFIWIGQINDTAINYINQWKDVS-----DYTVKVFDSNAF----- 136
 DB 758 QMGKEGLPTE-----EIKNVLEKVSSEWKNVSCHLEDLERKIQLOEDINAYPKQL 807
 QY 137 -----LINT-----LKKTIVSATNTLTSPRE-----NL----- 161
 DB 808 DELEKVIKTEEWVYKHTSISESSRQSLPSLKDSQORELTNLGLHPKIEMARASCSALMS 867
 QY 162 --NDPEFDYNKF-----YRGRMEIYDKQKHFDIYYSQ-----IENPFIIDNI 205
 DB 868 QPSADPFVQRGDFSLGRYQAVQEAVEDRQCHLENELKGPQHAYLETTLTKLDKVLNDSE 927
 QY 206 IKTVLSNEYSKDLEALNKYTEE--SINKITANN-----GNDIRNLEKFADEDLVLYN 256
 DB 928 NKAQVSLNVLNDLAKVEKALQEKTKLDEILENQKPAHLKLAEBETKALEKNVHPDVEKLYK 987
 QY 257 QELVE--RWN---LAAASDI-----LRISMUKEDGGV---YLDVDILPGIOPDLFKS 300
 DB 988 QEFDDVQGMKNKLVLSKDLHLEETALTURAEADSTVIEKWD-----GVKDFLMKQ 1042
 QY 301 -----INKPDSITNTSWEMIKLEAINK-----YKEYIP-----GYT 331
 DB 1043 QAAQDDAGLQRLDQCSAFVNEIETIESILKNMKEIETNLRSRGPVAGIKTWVQTRLGDY 1102
 QY 332 SKNFDMLDEEV-----QRSFESALSSKDSKSEIFLPDDIKVSPLEVKIAPANNVIN 384
 DB 1103 QTQLEKLSKETATQKSRLSSESQEAANLKOLAEMQEWMTQAEBEYLERDPEYKSPPELE 1162
 QY 385 QALISLKDSYCSDLVINQIKNRYKILANDNLNPSINEGTFNTWKIIFSDKLASISNEDNM 444
 DB 1163 SAVEEMKRAK-EDVLQKEV--RVKILKXNI-----KLAAKVPFSGGQELTS 1205
 QY 445 MFMTKITNYL-----KVGFAPDVRASTINLSRGPVVTGAYQDLMFKD--NSTNIHLE 495
 DB 1206 ELNVVLENYQLLCNRIRGKCHTLEVWS-----CWIELHYLDLETWLTLE 1253
 QY 496 PELRNFEPPKTKISQLTEQETISLWSP--NOARAKSQPEEYKGYFGALGEDNLDPAQ 553
 DB 1254 ERMKSTEVLPEKTDVANE-ALESLESVLRHPADNRTQIRELGQTLIDGG----- 1301
 QY 554 NTVLDKQVYKKSILSSMKTRNKEYIHIVQCGDKISYEACNLFPSKDPYSILYQKNIE 613
 DB 1302 --ILD-DIIESEK-LEAFNRSYEDLSHL---AESKQISLEKQLQVLRDQMLQVLESGL 1354
 QY 614 GSETAYYYYVADAEIKDKYRIPYQIS-----NKRNIKLFIGHGK 655
 DB 1355 ELQKQLTTLTLD---RIDAFQVPOEAKIQAEISAEHLTLEELRRNRSQPLT----- 1404
 QY 656 SEFNTDTFANLDVDSLSEIETILNLAADIAPKYIEINILGCMNFSYSISAEETYPGKL 715
 DB 1405 SPESRTARGGSGQMDVLQKLEEVSTKQFLQKQFANFQRMDCRVLDDGVKAE----- 1457
 QY 716 LLKIKDRVSELMPSISQDSITVSANQYEVRIINEGRKREILDHSGKWINKEESIIX---DI 772
 DB 1458 -LHVLD-VKDVPDVIQTHLDCKMKLYKTL--SEVKLEV-----ETVIKTRGHI 1502
 QY 773 SSKEYISPNKPKENKIIIVKSKYLHELSTLLQBEIRNANSSDIDLEKKWMLTECEINVASNI 832
 DB 1503 VQKQQTQD-NPKG-----MDEQLTSLKVLNDLGAQVTEGKQDLERASQARKMKKEASL 1556
 QY 833 DRQIVEGRIIEAKNLTSD-----SINYIKNFKELIE-----STSDSLYDLKHONG 877
 DB 1557 SEWJSATETELVQKSTSEGLGLDLDTEISWAKNVKOLEKPKAKDLNTITESAAL--QNL 1614
 QY 878 LDDSHFTSFEDISKTEGFI-RIRPINKETGNSIF-IETEKEIFSEYATHISKEISINIKDT 935
 DB 1615 IEGSEPILEERLCVNLNAGWSRVRTWEDWCNTLNMHQNQLEIFDGNVAHISTWLYQ-AEA 1673
 QY 936 IPDNVNGK-----LVYKVV--NLDAAH--EVNTLNSAFFIQS-----LIEY 971
 DB 1674 LLDETEKPTSKQBEIVKRLVSELDDANLQVENVRDQALILMNARGSSRELVEPKLAEL 1733

QY 972 NTKESLS-----NLSVAM-----KVQVYQALFTGL-----NTTIDASKVVB--L 1010
Db 1734 NRNEKVSQHIKSAKLLIAQPLVQCLVTTFTFTGTFVDFSDLEKLENDIENMLKFEVXHL 1793
QY 1011 VSTLDETDIDLLPTLSEGLPIIATIIDGVSLGAIAKSELSETNDPLLRQEE-E-----AKIGI 1066
Db 1794 ESSDEDERWD-----BESAQEEVLQRGEEMLPHPMEDNKKKIRL 1834
QY 1067 MAVNUTAASTAVTSAL-----GIASFSLIIVLPL-----AGISAGIPS 1105
Db 1835 QULLLHRYNKIKAIPIQQRKMGQASIRSLPFTDYLVEINKILLCMDDVLSLNVPE 1894
QY 1106 LVNNELIQQDKATKVIDYFKHSIAETEGATFLDDDKIIM---PQDDLVLSIEDFNNSI 1162
Db 1895 L--NTAIYEDFS-----FQEDSLKNHKQDQKLGEQIAVHEKQPDVIL----- 1936
QY 1163 TLGKEIWRAGSGG-----HTILT-----DDIDHFFSPSPITY---RKPMLSIYDVNLNIK 1209
Db 1937 -----EASGPEAIQIRDTILTQNAKWDRINRMYSDRKCFDRAEWEWRQFHCDDLNDL 1988
QY 1210 KEKIDFSKDLMLVLPNAPN-----RVFGYEM-----GWTPGFRSLNDGTLKLLDRD 1256
Db 1989 TQWITEABELLVDTCAPGSLDLEKARTHQELEVGISSHQPSAALNRTGDGIVQKLS 2047
QY 1257 HYEGOFYWRYPAFIA---DALITKLPYEDTVNVRINLDGNTRSFIVPVITTEQIRKNLS 1313
Db 2048 QADGSLFKLEKLAGLNQRDAIVAEYKDR-----QPLKGESK-----QVMK--- 2088
QY 1314 YSFYGGSGSYSLSPYNNIDNLVENDTWIDVDNVV-KNITIE-SDEIQKGELIENI 1371
Db 2089 -----YHQD-----EIIICWLTRAHAMQKRSITELGENLOE---LRDL 2125
QY 1372 LSKLMIEDNKI--ILNNHTINFGD-----INESNRFTISITFS--ILEDI-----NI 1414
Db 2126 TQSEMEVHAELKWLARTLEMLSDLSLPERDKISESLRTVNNWTKICREVPETTLKEC 2185
QY 1415 IIEIDLVSXKYLILSGCMK--LIENSSDI 1443
Db 2186 IQEPSSVSQT-RIAAHPNVQKVLVSSASDI 2215

RESULT 15
MLP1_YEAST
ID MLP1_YEAST STANDARD; PRT; 1875 AA.
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RP STRAIN=S288c;
RC MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
[2]
SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Ramacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UB12 and MLP1 genes and three
RT new open reading frames.";
RT Yeast 9:1349-1354(1993).
CC -!- FUNCTION: Myosin-like protein that is probably involved in DNA
CC repair.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.

CC -!- CAUTION: Ref.2 misquotes the gene name as "MPL1".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L01992; AAA34783.1; .
DR EMBL: X73541; CAAS1948.1; .
DR EMBL: Z28320; CA882174.1; .
DR PIR: S38173; S38173.
DR GeneOnline: L40074; .
DR SGD: S0001803; MUP1.
DR GO: GO:0005635; C:nuclear membrane; IDA.
DR GO: GO:0005654; C:nucleoplasm; IDA.
DR GO: GO:0006606; P:protein-nucleus import; IDA.
KW Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 3.3%; Score 288; DB 1; Length 1875;
Best Local Similarity 18.4%; Pred. No. 0.00097;
Matches 314; Conservative 304; Mismatches 612; Indels 474; Gaps 75;

QY 226 EESLNKITANNNGDNRNLEKFAEDLVRLYNOLV-----ERNWLAASDILRISMLK 278
Db 18 DERLNAIAFFQCSIEQVAKSF-DGVVVKHLNDKLLQFNLKSENKLVTVSPFELKASSLK 76
QY 279 EDGGVYLDVDDILPGIQPDLFKSIN---KPSDITNTSWEMI-----KLEAIMKYKEY 326
Db 77 KIDGLKTEMENIVRENDKIRKERNDTFFVKFSEVENEKMKLSSELEFVVKRKLDDLTEEKKE 136
QY 327 IPGYTSKNFMDLDEEVQ-----RSFESALSKSKSKSEIFLPLD-----D 365
Db 137 TQSNQORTLKLDERLKEIYELVRVNNRSNCKLRSTIMDLTKQQQYITNDLSRTE 196
QY 366 IKVSPLEVKIAFANNSVINQALISLKDYSYS-----DIVINOIKRYKIL----- 410
Db 197 LERKTOELTLLQSNNDWLEKELRSKNEQYLSVRQTKDVKILDIRNELNRLNDFQVETRN 256
QY 411 NDNLNPSINEGDFNTTMTKIFSDKLASIS-----NEDNMFMWIKITNYLKVGFAPDVS 464
Db 257 NDVLKORNE-----LSKSLQEKLEIKGLSDSLNSEKQEFSAEMS--LKQRLVDLLES 308
QY 465 TINLSGPGVYTGAYQDMLFMKDNSTNIHLE-----PE-----LRNFEFPKTKISQLTQ 514
Db 309 QLN-----AVKEELNSIRELNTAKVIADDSKKQTPENEDLLKELQLTKEKLAQ-CEK 359
QY 515 EITSLWSFNQARAKSOFEEYKKGYPFEGALGEDDNLDFQNTVLDKDYVSKKILSNKTRN 574
Db 360 ECLRLSSITDE-----ADEDNENLSAKSSSDFFLKKQIKERT-- 399
QY 575 KEVIH-----YIVQLQGDKI-----SYEASCNLSKDPYSSILYQKNIEGSEYAYYYVAD 625
Db 400 KEHLQNLQIETFTIVELE-HKVPIINSFKERTDMLNENLNNAALLLETSNEK-----N 450
QY 626 AEIKEDIKVRIPYQISGNKNIKLTIFGHGKSBNFTDTFANLDV-----DSLSS-- 673
Db 451 AKVKEL-----NAKQKLV-----ECENDLQITLTKQRLDCROIQYLLITNSVNSDS 497
QY 674 -----EIEITLNLAADISPKYIEINLLGNMFSYSASSETYPGKLLKIKRVSSEL 726
Db 498 KGPLRKEEIIQIQNIMQEDDSTI-----TESDSQKVTVTERLVEFKNIQIQEKNAEL 549
QY 727 MPISQSDSITVSANQYEVRAINEREGKEIILDHGSKWINKESIKIDISSKEYISFNPKENK 786
Db 550 LKVVNR-----LADKLESK-EKKSQSLQKIBSETVNEAKEAITLTKSRM-----DLERS 599

Search completed: April 1, 2004, 16:42:35
Job time : 35.5 secs

QY 787 IIVKSYLHSLTLLQEIIRNANSDDILEKKVMLTECEINVASNIDROIVGRIEBAKN 846
Db 600 IEBLQKLEBELTSP--NEDASYNVTKQ---LTETKRDLESQV--ODLQTRISQITR 652
QY 847 LTSDSINYIKNEPKLIESDSLYDLKHQGLDDSHFISFE-----DISKTENG- 895
Db 653 ESTENMSLLNKE---IQDLYDSKSDISIKLGEKSSRIILAEERFKLLSNTLDTLTKAENDQ 709
QY 896 FRIRFINKGTGNSFIETEK--EIFSEYATHISK-----EISNKOTIFDNNKGLVKX 947
Db 710 LRKRPDYLO--NTILKQDSKTHETLNEYVYCKSLISIVETELLNLKE-----EQKLRVH 761
QY 948 VNLDAHEVNTLNSAFFIQLSLEYNTKESLNSLVAMK-----VQVYAQLFSTGINTIT 1002
Db 762 LEKMLKQELNKLSP-----EKDSLRTVTLQTLQKEREDELLSETRKSCQKIDELE 813
QY 1003 DASKVVELSTALDETI-----DLLPTLSEGLPIIATIIDG----- 1038
Db 814 DALSELKETSQKDHKIQLEEDNNSNIWYQNKIEALKDYESVITSVDSKQTDIEKIQ 873
QY 1039 ---VSLGAIKE-----LSET--NDPLLRQEIEAKIGIMAVNLTAATAI----- 1078
Db 874 YKVKSLKEFEEDKIRLHTVNVWDETINDSLRKELEKS---KINLTDAYSQIKEYKDL 929
QY 1079 --VTSALGIASGPSILLVPLAGISAGPSLVNNELLQDKATKVIDYFKHISLAETEGAF 1136
Db 930 YETTSQ-----SLQOTNSKLDSEFKDF---TNQTK 956
QY 1137 TLDDDKLIMPDDLVLSEIDFN--NNSITLCKEIMWAEAGSGHGLTDDIDHFFSSPSITY 1195
Db 957 NLTEKTSLEDKISLLKEQMFNLNNELDIQ----- 987
QY 1196 RKPWLSIYDVNLNKKIKPFSKDLVLPNAPNRFVGYEMGTFPGFSLDN----- 1245
Db 988 -----KGMEKAKAFKXISILQNNKEVEAVKSEYSEKSLKIQNDLDQOTIYAN 1037
QY 1246 -----DGTKLDRDRH---YEGQFYWYFAPFADALITKLKPYED--- 1284
Db 1038 TAQNNYEQELQHEADVSKTISELREQLHTYKGV--KTENLSRDQLENALKENERSWSQ 1095
QY 1285 -----TNVRI--NLGNTSPFVITTEQIRKNLSYSPYSGGYSLSLFPYNN 1333
Db 1096 KESILLEQLDLSNGRIEDLSQNKLYDQIYTAADKEVNNSTNGFLNNILITLRRERD 1155
QY 1334 I-----DLNLVENDTWJVDNVNKNITIESDEIQ--GELI---ENILSKLN- 1376
Db 1156 ILDTKVTVARDAKMLRQKISLMDVQLQARTKLDNSRVERKENHSSIIOOHDDIMEKLNQ 1215
QY 1377 ---IEDNKIILNHTINFYGDINESNRFTSLFSEILEDINIIIEIDLVSYSKILLSGNC 1433
Db 1216 LNLRESNITLRNLEN-----NNKKELQSELDKLNQNVAPIESELTALKYSMQEKEQB 1271
QY 1434 MKLI-----ENSSDIQKIDHIGFNGEHQKIPYYSIDN-----ETK 1470
Db 1272 LKLAKERVHWKKSQDILEKHQLS--SSDYKL--ESBIENLKELENKERQGAABEK 1328
QY 1471 YNGFIDYSKEGLFTAEFNSIESIRNI--YMPDSNNLFIYSSKDLKDIRINKGDVKLLIG 1529
Db 1329 FNR--LRQAQERLKTSLQCSLTEQVNSLRDARV--LENSLSEANARIEELQNAKVAQG 1386
QY 1530 N-----YFKDDMKYS--LSFTIEDTNT---IKNGVYLDENGVAQILKFWNAKSAL 1576
Db 1387 NNQLEAIRKLOEABAKASRELOAKLEBESTSYSTINGLNEEITTLKEEIEKORQIOQOL 1446
QY 1577 -----NTSLSMNFLESI-----NIKNIFFYNLDPNPIEFLDTNFIISGNSIQFEL 1624
Db 1447 QATSANEQNDLSNIVESMKKSFEEDKIKFKIKEXTQEVN--EKILEAQERLNLQPSNI--NMEE 1504
QY 1625 ICDKKNIQPYFNFKIKETS YTL 1648
Db 1505 IKKKWESEHEQVSKIRAEAL 1528

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:38:25 ; Search time 23.75 Seconds
(without alignments)
6895.288 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700

Perfect score: 8675

Sequence: 1 MNLVNAQLQKVVYKFRIO.....VLYGIDRYVNVKVIAPNLYT 1700

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8675	100.0	2364	2 I40884	cytotoxin L - Clos
2	6759.5	77.9	2366	2 S10317	toxin B - Clostrid
3	6737	77.7	2367	2 S70172	toxin A - Clostrid
4	4467	51.5	2710	2 A37052	alpha-toxin - Clos
5	2120.5	24.4	2178	2 S5805	toxin B - Escheric
6	507.5	5.9	3169	2 T00296	rhostry protein -
7	468.5	5.4	2401	2 T28676	rhostry protein -
8	440	5.1	2269	2 T28677	hypothetical prote
9	431.5	5.0	4688	2 F62885	hypothetical prote
10	422	4.9	2166	2 G70163	hypothetical prote
11	399	4.6	3724	2 T18427	hypothetical prote
12	398	4.6	1979	2 C71622	hypothetical prote
13	391	4.5	3225	2 D81702	adherence factor T
14	382.5	4.4	3394	2 T18501	hypothetical prote
15	368	4.2	1711	2 T18429	hypothetical prote
16	361.5	4.2	1939	2 T18372	repeat organelle
17	360	4.1	2829	2 A42771	reticulocyte-bind
18	348	4.0	1127	2 T28317	ORF MSV156 hypothe
19	345.5	4.0	4981	2 T18489	hypothetical prote
20	344	4.0	2819	2 A90551	conserved hypothet
21	343	4.0	2485	1 H71621	serine/threonine-s
22	340.5	3.9	5005	2 F62884	hypothetical prote
23	338.5	3.9	3335	2 H81702	adherence factor T
24	331	3.8	3255	2 G81702	adherence factor T
25	329.5	3.8	4550	2 T18440	hypothetical prote
26	329	3.8	2910	2 T28156	DNA-directed RNA p
27	323	3.7	1819	2 D97033	uncharacterized pr
28	318.5	3.7	3973	2 B71612	hypothetical prote
29	317.5	3.7	2599	2 F90608	ABC transporter pe

30 315 3.6 1302 1 JC6009 surface-located me
31 313.5 3.6 3844 2 T18402 asparagine/asparta
32 312 3.6 1639 2 S05603 major merozoite su
33 311.5 3.6 1726 1 SAZQGM major merozoite su
34 311.5 3.6 1726 2 A45948 major merozoite su
35 311 3.6 1516 2 E71619 RAD2 endonuclease
36 310 3.6 1306 2 T28313 ORF MSV152 probabl
37 309.5 3.6 1447 2 F82909 hypothetical prote
38 309.5 3.6 1701 2 A28688 major merozoite su
39 309.5 3.6 1957 2 T38077 hypothetical coile
40 309 3.6 2244 2 F90563 hypothetical prote
41 307.5 3.5 1701 2 A54498 major merozoite su
42 307.5 3.5 1830 2 E82909 conserved hypothet
43 305.5 3.5 2496 2 A71616 secreted protein p
44 304.5 3.5 1802 2 G71616 hypothetical prote
45 302.5 3.5 1252 2 B42771 reticulocyte-Bindi

ALIGNMENTS

RESULT 1

I40884

Cytotoxin L - Clostridium sordellii

C;Species: Clostridium sordellii

C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999

C;Accession: I40884

R;Green, G.A.; Schue, V.; Montell, H.

Gene 161, 57-61, 1995

A;Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium

A;Reference number: I40884; MUID:95369733; PMID:7642137

A;Accession: I40884

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-2364 <RES>

A;Cross-references: EMBL:X82638; NID:g1000694; PIDN:CAAS7959.1; PID:g1000695

C;Superfamily: cpl repeat homology

C;Keywords: cytotoxin

Query Match 100.0%; Score 8675; DB 2; Length 2364;

Best Local Similarity 100.0%; Pred. No. 2.4e-304; Indels 0; Gaps 0;
Matches 1700; Conservative 0; Mismatches 0;

QY	1	MNLVNAQLQKVVYKFRIOEDEYVAILNALBEYHNMSSESVVEKYLKLDINLNDNYL	60
DB	1	MNLVNAQLQKVVYKFRIOEDEYVAILNALBEYHNMSSESVVEKYLKLDINLNDNYL	60
QY	61	NTYKSGRNKALKKFKYELTMEVLELKNNSLTPVEKNLHFIIWGGQINDTAINYINQWKD	120
DB	61	NTYKSGRNKALKKFKYELTMEVLELKNNSLTPVEKNLHFIIWGGQINDTAINYINQWKD	120
QY	121	VNSDYTVKVFYDSNAFLINTLKKTIIVESATNTLTESFRENLDNDFDYNNKFKRMEIYY	180
DB	121	VNSDYTVKVFYDSNAFLINTLKKTIIVESATNTLTESFRENLDNDFDYNNKFKRMEIYY	180
QY	181	DKOKHFDYKSGQIEENPEFIIDNIKTYSNEYSKDLKALNKYIEESLNKITTANNNDI	240
DB	181	DKOKHFDYKSGQIEENPEFIIDNIKTYSNEYSKDLKALNKYIEESLNKITTANNNDI	240
QY	241	RULEKFADEDLVRLVNLQELVERNLAAASDIIRISMLKEDGGVYLDVDILPGIQDLPFKS	300
DB	241	RULEKFADEDLVRLVNLQELVERNLAAASDIIRISMLKEDGGVYLDVDILPGIQDLPFKS	300
QY	301	INKPDSITNTSWEMIKLEAIKMKYKIPGYTSKNFMDLDEEVQSFESALSSKSKSEIF	360
DB	301	INKPDSITNTSWEMIKLEAIKMKYKIPGYTSKNFMDLDEEVQSFESALSSKSKSEIF	360
QY	361	LPDDDKVSPLEVKIAFANNSVINQALISLKSYSDDLVIINQIKRYKILNNDLNPISNE	420
DB	361	LPDDDKVSPLEVKIAFANNSVINQALISLKSYSDDLVIINQIKRYKILNNDLNPISNE	420
QY	421	GTDFNTMTKIPDKLASISNEDNMFMIKITNLYKVGAPDVRSTINLSGPGVYTCAQD	480

Db 421 GTDFNTMK:FSKLASISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTCAQD 480
QY 481 LLAFKONS:TNHLEPELRNFPFKTKISOLTEQETISLWSNQARAKSOFEEYKKGYPE 540
Db 481 LLAFKONS:TNHLEPELRNFPFKTKISOLTEQETISLWSNQARAKSOFEEYKKGYPE 540
QY 541 GALGEDNDLFAQNTVLDKDYVSKILSSMKTRNKYIHYIYVLOQDKISYEASCNLFPSK 600
Db 541 GALGEDNDLFAQNTVLDKDYVSKILSSMKTRNKYIHYIYVLOQDKISYEASCNLFPSK 600
QY 601 DPYSSILYQKNIEGSETAYYYVADABEIKEDKIRIPYOISNKRNIKLTFFIGHGKSEFNT 660
Db 601 DPYSSILYQKNIEGSETAYYYVADABEIKEDKIRIPYOISNKRNIKLTFFIGHGKSEFNT 660
QY 661 DTFANLDVDSLSSEIETILNLAADISPKYIRINILGCMNFYS:SAEETYPCKLLLIK 720
Db 661 DTFANLDVDSLSSEIETILNLAADISPKYIRINILGCMNFYS:SAEETYPCKLLLIK 720
QY 721 DRYSELMPISQDSITVSANQEVNRINEGKEIILDSKWKINKESIIKDISKEYISF 780
Db 721 DRYSELMPISQDSITVSANQEVNRINEGKEIILDSKWKINKESIIKDISKEYISF 780
QY 781 NPKENKIIYKSKYLHELSTLLQEIRNANSDDILEKKVMLTECEINVASNIDRQIVEGR 840
Db 781 NPKENKIIYKSKYLHELSTLLQEIRNANSDDILEKKVMLTECEINVASNIDRQIVEGR 840
QY 841 IBEAKNLTSDSINYIKNEPKLESISDSLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
Db 841 IBEAKNLTSDSINYIKNEPKLESISDSLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
QY 901 INKETGNSFIETEKIPEYATHISKESINIKOTIFDNVNGKLVKYNLDAAEHVNTLN 960
Db 901 INKETGNSFIETEKIPEYATHISKESINIKOTIFDNVNGKLVKYNLDAAEHVNTLN 960
QY 961 SAFFIQSLIYNTKESLSNLSVAMKVQVYAOLFSTGLNTITDASKVVELYSTALDETID 1020
Db 961 SAFFIQSLIYNTKESLSNLSVAMKVQVYAOLFSTGLNTITDASKVVELYSTALDETID 1020
QY 1021 LLPTLSEGLPIIATIDGVSGLAAIKELSETNDPLLROEIEBAKIGIMAVNLTAATAIYT 1080
Db 1021 LLPTLSEGLPIIATIDGVSGLAAIKELSETNDPLLROEIEBAKIGIMAVNLTAATAIYT 1080
QY 1081 SALGIASGSIILVPLAGISAGIPSLVNNELILOKATKVLDYFKHISLAETGAFTLLD 1140
Db 1081 SALGIASGSIILVPLAGISAGIPSLVNNELILOKATKVLDYFKHISLAETGAFTLLD 1140
QY 1141 DKIIIMPQDDLVLSEIDFNNSNITLCKEIBWRAEGSGHLLTDDIDHFFSPSSIYIRKPEWL 1200
Db 1141 DKIIIMPQDDLVLSEIDFNNSNITLCKEIBWRAEGSGHLLTDDIDHFFSPSSIYIRKPEWL 1200
QY 1201 SIYDVLNKEKIDFSDKLMVLPNAPNRFVGYEMGWTGFRSLNDGDKLLDRDRHYEG 1260
Db 1201 SIYDVLNKEKIDFSDKLMVLPNAPNRFVGYEMGWTGFRSLNDGDKLLDRDRHYEG 1260
QY 1261 QYWYFAPFIADALITKLPYEDTNVRINLDGNTRSPFIVPITTEQIRKNLSYSFYGS 1320
Db 1261 QYWYFAPFIADALITKLPYEDTNVRINLDGNTRSPFIVPITTEQIRKNLSYSFYGS 1320
QY 1321 GSYLSLSPYNNIDNLVENDTWIVDYNVVKNTIIESDEIQGELIENILSKLNTEDN 1380
Db 1321 GSYLSLSPYNNIDNLVENDTWIVDYNVVKNTIIESDEIQGELIENILSKLNTEDN 1380
QY 1381 KIILNNHTINFGDINESNRISLTFSTLEIDNIIIEIDLVSQYKILLSGNCMKLIENS 1440
Db 1381 KIILNNHTINFGDINESNRISLTFSTLEIDNIIIEIDLVSQYKILLSGNCMKLIENS 1440
QY 1441 SDIOQKIDHIFGNGHQYIIPYSYIDNETKNGYFIDYSKKEGLFTAFSINESIIRNYMP 1500
Db 1441 SDIOQKIDHIFGNGHQYIIPYSYIDNETKNGYFIDYSKKEGLFTAFSINESIIRNYMP 1500
QY 1501 DSNNLFIYSSKDLKDIRINKGDVKLLIGNYFKDDMKVSLFTTEDNTIKNGVYLDEN 1560
Db 1501 DSNNLFIYSSKDLKDIRINKGDVKLLIGNYFKDDMKVSLFTTEDNTIKNGVYLDEN 1560

RESULT 2

S10317

toxin B - Clostridium difficile

C:Species: Clostridium difficile

C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 15-Oct-1999

C:Accession: S10317; S21894; S22434

R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.

Nucleic Acids Res. 18, 4004, 1990

A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.

A:Reference number: S10317; MUID:90326540; PMID:2374729

A:Accession: S10317

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-2366 <BAR>

R:Michel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn,

submitted to the EMBL Data Library, July 1991

A:Title: Comparative analysis of Clostridium difficile toxins A and B.

A:Reference number: S21894

A:Accession: S21894

A:Molecule type: DNA

A:Residues: 1271-2366 <EIC>

A:Cross-references: EMBL:X60984; NID:G40445; PIDN:CAA43299.1; PID:G40446

Mol. Gen. Genet. 233, 260-268, 1992

A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.

A:Reference number: S22434; MUID:92293124; PMID:1603068

A:Accession: S22434

A:Molecule type: DNA

A:Residues: 1791-2366 <VON>

A:Cross-references: EMBL:X60984

C:Genetics:

A:Gene: toxB

C:Superfamily: cpl repeat homology

C:Keywords: cytotoxin

Query Match 77.9%; Score 6759.5; DB 2; Length 2366;

Best Local Similarity 76.6%; Pred. No. 1.9e-235;

Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;

QY 1 MNLVNAQLOKVMYKVFRIQDEYVAILNALUEYHNMSSESVVEKYLKDKINLTDNYL 60

Db 1 MSLVNAKLEKMANVRFQDEYVAILDALEYNMSNTVVEKYLKDKINLTDYI 60

QY 61 NTYKSGRNKALKPKFVLTMEVLEKKNLSLTPVEKNLHFVIGGQINDTAINYNQWKD 120

Db 61 DTYKSGRNKALKPKFVLTVEVLEKKNLTPVEKNLHFVIGGQINDTAINYNQWKD 120

QY 121 VNSDYTVKVFYOSNAFLINTLKTIVESATNTLSPRENLDPEFDYKFKMEIY 180

Db 121 VNSDYVNVVYFOSNAFLINTLKTIVESALNDTLESFRENLDPRFDYKFKMEIY 180

QY 181 DKQKFIIDYKQIOBENPEFIIDNIIKTVLSNEYSKOLEALNKYTEESINKITANNNDI 240

Db 181 DKQKFIIDYKQIOBENPEFIIDNIIKTVLSNEYSKOLEALNKYTEESINKITANNNDI 240

QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDIIRISMLKXEDGGVLDVDILPGIQDLPFS 300

Db 241 RNLEKFADEDLVRLYNQELVERWNLAAASDIIRISMLKXEDGGVLDVDILPGIQDLPFS 300

A:Molecule type: DNA	
A:Residues: 1-55 'V', 57-2079 'L', 2081-2549 'S', 2551-2710 'SAU'>	
A:Cross-references: EMBL:X51797	
C:Genetics:	
A:Gene: toxA	
C:Superfamily: Clostridium difficile toxin A; cpl repeat homology	
C:Keywords: cytotoxin; enterotoxin	
F:1820-1839/Domain: cpl repeat homology <CP01>	
F:1840-1860/Domain: cpl repeat homology <CP02>	
F:1861-1881/Domain: cpl repeat homology <CP03>	
F:1933-1952/Domain: cpl repeat homology <CP04>	
F:1953-1973/Domain: cpl repeat homology <CP05>	
F:1974-1994/Domain: cpl repeat homology <CP06>	
F:1995-2015/Domain: cpl repeat homology <CP07>	
F:2067-2086/Domain: cpl repeat homology <CP08>	
F:2087-2107/Domain: cpl repeat homology <CP09>	
F:2108-2128/Domain: cpl repeat homology <CP10>	
F:2129-2149/Domain: cpl repeat homology <CP11>	
F:2201-2220/Domain: cpl repeat homology <CP12>	
F:2221-2241/Domain: cpl repeat homology <CP13>	
F:2242-2261/Domain: cpl repeat homology <CP14>	
F:2315-2334/Domain: cpl repeat homology <CP15>	
F:2335-2354/Domain: cpl repeat homology <CP16>	
F:2356-2376/Domain: cpl repeat homology <CP17>	
F:2377-2397/Domain: cpl repeat homology <CP18>	
F:2449-2468/Domain: cpl repeat homology <CP19>	
F:2469-2489/Domain: cpl repeat homology <CP20>	
F:2490-2510/Domain: cpl repeat homology <CP21>	
F:2562-2581/Domain: cpl repeat homology <CP22>	
F:2582-2602/Domain: cpl repeat homology <CP23>	
F:2653-2672/Domain: cpl repeat homology <CP24>	
F:2673-2694/Domain: cpl repeat homology <CP25>	
Query Match 51.5%; Score 4467; DB 2; Length 2710;	
Best Local Similarity 50.7%; Pred. No. 6 2e-153;	
Matches 87; Conservative 366; Mismatches 440; Indels 42; Gaps 15;	
Qy	1 MNLVKAQKQVYVVFRIODEVAILNALBHYHNMSSVYKYLKLDINLTDNYL 60
Db	1 MSLISKELIKLAY-SIRPRENEKYTLTILNLDYVNLKLTNNENKYLQKLKLNESIDVFM 59
Qy	61 NTYKSGRNKALKKFEYLTVNLELKNLSLTPVEKNLFIWIGGQINDTAINYNQWKD 120
Db	60 NKYTSSRNALNLKDKILKILKILKILKILKILKILKILKILKILKILKILKILK 119
Qy	121 VNSDYTVKVFVDSNAFLNLTAKTIVESATNTLESFRENLDPEFDYKFKRMEITY 180
Db	120 INAEYNKILWYDSEAFVNLTKKALVESSTTEALQLEBEIQPNQFDNMKFKRMEITY 179
Qy	181 DKQGHFDYKSGTEENPEFIQNIKTYSNEVSKOLEALNKYIEESLNKIPANNNDI 240
Db	180 DRQKRFYNYKSGQINKFTVFTIDDIKSHLVSEYNRDETVLESYRTNSLRKINSNGIDI 239
Qy	241 RNLEKFADEDLVRLYNQELVERNLAAASDILRSMLEKDEGGYLVLDLPGIQPLFKS 300
Db	240 RANSLFTEQELLNYSQELLNLRGNAASAADIVLLALKNFGVYLVLDVLMPLGHISDLFT 299
Qy	301 INKPDSTNTSWEMIKLEATMKYKHYIPGYTSKNFMDLDEEVORSFESALSSKSDSEIF 360
Db	300 ISRPSSIGLDRWEMIKLEATMKYKHYINNTSENFDKLDQDLNFKLIESKSESEIF 359
Qy	361 LPDDIDKVPLEVKIAPANNNSNOALISLKDCSDGLVINOIKNRYKILNLDNPSINE 420
Db	360 SKUENLVSDLEIKIAFALGSVINOALISKQSGSLVNLVTEQVKNRYQFVNOHLNPAIES 419
Qy	421 GTDFNTMTKIFSDKLASINEDNMFMKIKITNLKVGFPADPVRSTNLSGPGVYTGAYQD 480
Db	420 DNNFTDTTKIFHDSLFNSAFAENSMFTKTAPLYQVGFPEARSTLSGPGAYASAYD 479
Qy	481 LLMPKDNSTNHLPELRNPEFPKTKISQLTQETISLWSFNQARAKQFEBYKGYPE 540
Db	480 FINLQENTIEKTLKASDLIEFKFPENNLSQLTEQINLSLWSPQASAKYQFEBYVRDYG 539
Qy	541 GAGEDNDLPAQNTVLDKDY-VSKILSS--MKTRNKYIHYIVQLQGGKISYEACNL 597
Db	540 GSLSEDGVDNKATLADKXVLLNNKIPSNVNEBAGSKYVHYIIQOGDDISEATCNL 599
Qy	598 FSKDPYSILYQKNIEGSETAYYYVAD--AEIKEDKIRIPYQISNKRNIKLTFIGHGK 655
Db	600 FSKNPKSIIIRNN--NESAKSYFLSDGSESILELNKIRIPERLKNKEKVKVTFIGHGK 657
Qy	656 SEFNTDTPANLDVSLSEIETILNADKADISPKYIEINLLGCMNMFYSISAEITYFGKL 715
Db	658 DEFNTSEFARLSVDSLSEISSFLDTIKDISPNVENVLLGCMNMFYSYDFNVEITYFGKL 717
Qy	716 LLKIKDRVSELMPSISDSITVSANOYEVRINEREGKEILDHSGKWINKESIIKDISSK 775
Db	718 LLSIMDKITSLPDVKNKSITIGANQYEVINSEGRKELLAHSGWINKELAIMSDISSK 777
Qy	776 EYISFNPKNKIIVKSKYLHELSTLLOEIRNANSDDILEKVMLECEINVASNIDRQ 835
Db	778 EYIFDSIDNKLKAKSKNIPGLASISEDIKTLLLDASVSPDTKTLNANLKNIESSIGDY 837
Qy	836 IVEGRIEAKNLSDSINYIKNEFKLIESDSYDLAKHQGLDDSHFISPEDISKYENG 895
Db	838 IYVEKLEPVKNIIEHSDIDLDEFNLENVSDLEYELKLNLDKYLISPEDISKXNST 897
Qy	896 FRIRFINKETGNSIFIEKEIFSEYATHISKEISNKTOTIFDNVNGKLVKKNVLDAAHE 955
Db	898 YSVRFINKSGESYVETEKEIFSKYSEHTKETSITKNSIITVNGNLLDNIOQDHTSQ 957
Qy	956 VNTLSAPFIOSLIYNTKESLSNLSVAMKVOYVYAOQLFSTGLNTITDASKVSLVSTAL 1015
Db	958 VNTLNAAPFIOSLIDYSGNKVDLNDLSTSVKQYVYAOQLFSTGLNTIYDSIQVNLISAV 1017
Qy	1016 DETIDLLPTLSEGLPIIATIIDGVSGLAAIKELSETNDPLLRQRIEAKIGIMAYNLTAAS 1075
Db	1018 NDTINVLPTITEGPIVSTILDGINLGAAIKELADEHDLKKELEAKVGLAINMSLSI 1077
Qy	1076 TAVTSALGASGSIILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISAEATEGA 1135
Db	1078 AATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVWVNFHLSSEKYG 1137
Qy	1136 FTLDDDKIIMQDDLVLSLEIDFNNSITLKGCEIWRABGGSGHTLTDIDHFPSSPSITY 1195
Db	1138 LKTEDDKLVLPIDLVLSLEIDFNNSIKLGTCLNLAEGSGGHTVTGNIDHFPSSPSISS 1197
Qy	1196 RKPMLSIVDLNINIKKEIDFSKDLMLPNAPNRFVGYEMWTPGFRSLDNDGTLLDRIR 1255
Db	1198 HPSLSIYSALGISETENLDFSKIMLPNAPRPFVFWETGAVPGRLSENDGTLLDSIR 1257
Qy	1256 DHYEGQFYRYFAFIADALITKLPYEDTNRINLPGNTRSFIVPVITTEQIRKNSYS 1315
Db	1258 DLYPGKFYRYFAFF-DYAITLKPVEDTNIKIKLKDITENFIMPITITNEIRKLSYS 1316
Qy	1316 FYGGGYSLSLSPYNNMIDNLVENDTWIDVDNVVKNITIESEDETOGELIENILSKL 1375
Db	1317 FDGAGGYSLLSLSPYNTINLSKDDLWIFNIDNEVREISIENTGKKGKLIKDLVLSKI 1376
Qy	1376 NIEENKILNHNHTINFYGDINESNRFLSIFSLIEDNIIIEIDLVSYSKYLKSGCMK 1435
Db	1377 DINKNKLIIGNQITDFSGDIDNKORYIFLTCELDKDKSLIIEINLVAKSYSLLSGKNY 1436
Qy	1436 LIENSSDIDQOIKIDHGFNGEHQKYPISYID-NETKYNGFIDYSKKEGLFAEBSNESII 1494
Db	1437 LISLNSNTIEKINLGLD---SKNIAYNTDESNNKYFGAI-----SKTSQKSII 1483
Qy	1495 RNIIYDPSNNL-----FIYSSKOL--KDIRINKGDVKVLLIGNYFKO---DMKVSL 1541
Db	1494 H--YKDKSKNILEFVNDSTLEFNSKDFAEIDINVMKDDINTITGKYVDVNDNTDKSIDFS 1541
Qy	1542 FTIEDNTIKLNGVYLDENGVAQILKFNNAKSAINTSNLSMNPLESINIKNIYNNLDP 1601
Db	1542 ISLVSKNQVNGLYLAESVYSYLDYFVNSDGHNTSNFMNLFDINISFWKJGFGE--- 1598
Qy	1602 NIEFILDNTFIISGNSITGQELICDKNIQPYFINFKIKETSYTLVVGNRQNLIVEPS 1661